Efficient Control of Access Strata in LTE

by

Mohammad Tawhid Kawser

A Thesis Submitted to the Board of Examiners in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY IN ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT)
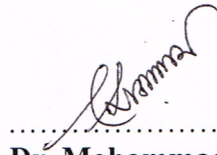Gazipur, Bangladesh

March 2016

Recommendation of the Board of Examiners

The thesis titled **"Efficient Control of Access Strata in LTE"** submitted by **Mohammad Tawhid Kawser**, student number 112702 of Academic Year 2011-2012, has been found as satisfactory and accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Electronic Engineering on March 28, 2016.
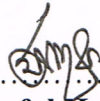
Board of Examiners:

1.

....
Dr. Mohammad Rakibul Islam   Chairman
Professor,   (Supervisor)
Department of Electrical and Electronic Engineering,
Islamic University of Technology (IUT), Boardbazar, Gazipur-1704.

2.

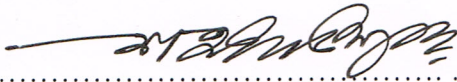..
Dr. Md. Ashraful Hoque   Member
Professor and Head,   (Ex-Officio)
Department of Electrical and Electronic Engineering,
Islamic University of Technology (IUT), Boardbazar, Gazipur -1704.

3.

....
Dr. Md. Shahid Ullah   Member
Professor,
Department of Electrical and Electronic Engineering,
Islamic University of Technology (IUT), Boardbazar, Gazipur -1704.

4.

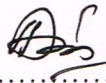..
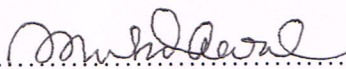**Dr. Md. Ruhul Amin**   Member
Professor,
Department of Electrical and Electronic Engineering,
Islamic University of Technology (IUT), Boardbazar, Gazipur -1704.

5.

....
Dr. M Abdul Awal   Member
Professor,   (External)
Department of Electrical and Computer Engineering,
North South University (NSU), Bashundhara, Dhaka- 1229.

i

Declaration of Candidate

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any Degree or Diploma.

Dr. Mohammad Rakibul Islam
Supervisor and Professor,
Department of Electrical and Electronic Engineering,
Islamic University of Technology (IUT),
Boardbazar, Gazipur-1704.
Date: March 28, 2016
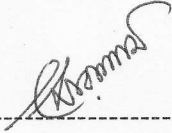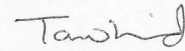
Mohammad Tawhid Kawser
Student No. 112702
Academic Year: 2011-2012
Date: March 28, 2016

Table of Contents

N.B. A CD containing the whole thesis is provided in the pocket of the back cover.

List of Tables

List of Figures

List of Acronyms

AM	Acknowledged Mode
APN	Access Point Name Packet Data Network PDN
ARQ	Automatic Repeat Request
AS	Access Strata
BER	Bit Error Rate
CSG	Closed Subscriber Group
DRX	Discontinuous Reception
EPC	Evolved Packet Core
EPS	Evolved Packet System
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
FFR	Fractional Frequency Reuse
GTP	GPRS Tunneling Protocol
HetNets	Heterogeneous Networks
HF	Handover Failure
HSS	Home Subscription Server
ICI	Inter-Carrier Interference
ICI	Inter-Cell Interference
ICIC	Inter-Cell Interference Coordination
IE	Information Element
KPI	Key Performance Indicator
LTE	Long Term Evolution
MAC	Medium Access Control
MBMS	Multimedia Broadcast Multicast Service
MME	Mobility Management Entity
NAS	Non-Access Stratum
NRT	Nonreal-Time
PBCH	Physical Broadcast Channel
PCC	Policy and Charging Control
PCCH	Common Control Channel
PCEF	Policy Control Enforcement Function
PCH	Paging Channel

PCI	Physical Cell Identifier
PCRF	Policy Control and Charging Rules Function
PDCP	Packet Data Convergence Protocol
PELR	Packet Error Loss Rate
P-GW	PDN Gateway
QoS	Quality of Service
RLC	Radio Link Control
RLF	Radio Link Failure
RRC	Radio Resource Control
RT	Real-Time
RTT	Round Trip Time
S-GW	Serving Gateway
SINR	Signal to Interference plus Noise Ratio
UE	User Equipment
UM	Unacknowledged Mode

Acknowledgements

Mohammad Tawhid Kawser

Abstract

The Long Term Evolution (LTE) is the latest Technology in the advancing series of cellular communication. The major key performance indicators (KPIs) of LTE are largely impacted by the access strata (AS), which handles the functionalities operating between the user equipment (UE) and the radio access network. This dissertation focuses on the achievement of better control of AS within a few important aspects while the operation of AS is broad. These aspects include power saving with discontinuous reception (DRX) operation, proper timing of cell switching for speedy users in heterogeneous networks (HetNets) and proper resource allocation and sectorization in HetNets. A few ideas have been put forward with a view to attain better control of AS in the respective areas.

The discontinuous reception (DRX) operation is included in LTE to achieve power saving and prolonged battery life of the UE. The DRX runs under the continuous control of AS. An improvement in DRX power saving usually leads to a potential increase in the packet delay. An optimum DRX configuration depends on the current data traffic, which is not easy to estimate accurately, especially, for nonreal-time applications. A particular fashion of variation in the DRX cycle length is proposed, which avoids continuous estimation of the data traffic when only nonreal-time applications are running with no real-time applications active. Since small delay in nonreal-time traffic does not essentially impact the user's experience adversely, a limited amount of delay is deliberately allowed in the proposal to attain significant improvement in power saving. The proposal also attempts to improve the delay in service resumption after a long inactivity. An analytical model has been established that performs stochastic analysis for both existing and proposed DRX methods. The model derives expressions for mean power consumption during data transfer, during wait period for packet scheduling and during the period of DRX cycles. The average power consumption is computed in a new way considering all three aforementioned power consumptions. Similarly, the model derives expressions for mean packet delay. MATLAB was used to simulate the analytical model with various assumptions in the model parameters. The power consumption and delay performances have been compared between the existing and proposed methods. It was observed that the power consumption significantly reduces in the proposed method while keeping the delay within acceptable limit. It was also observed that the proposed method significantly improves the delay in service resumption after a long inactivity.

The heterogeneous networks (HetNets) are a promising solution in LTE for ubiquitous and cost effective broadband user experience. But there are challenges to support seamless mobility in HetNets, especially when the user speed is high. These challenges have been investigated thoroughly and logical analyses have been performed to overcome them for the improvement of cell edge performance. The study indicates that AS should rigorously control the adaptation of cell switching parameters based on variation in coverage areas of different cells, traffic loads, user speeds, etc. A scheme has been proposed to scale cell switching parameters that incorporates Doppler spread based velocity estimation and adapts smoothly to various changes. The proposed method employs high controllability from AS. MATLAB was used to simulate and model the existing and proposed methods with various assumptions in the model parameters. It was observed that the lowest received power improves significantly and it becomes more stable with the variation in user speed in the proposed method. The improvement in several other KPIs was logically explained. However, the proposed method adds computational load and increases signaling overhead to some extent.

The Fractional Frequency Reuse (FFR) is a resource allocation technique that can effectively mitigate interferences in HetNets and it is a promising solution. Various FFR schemes have been suggested to address the challenge of interference in HetNets. The scopes of interference mitigation and capacity improvement have been studied. It is proposed that AS uses a resource allocation scheme that gradually varies frequency resource share with distance from the base station for both large and small cells in order to attain better utilization of the resources. This is performed effectively using three layers in the cell. The proposal also employs high number sectors in a cell, low interference and good frequency reuse. Monte-Carlo simulations have been performed using MATLAB and it was observed that the proposed scheme achieves significantly better throughput compared to the existing schemes.

It is expected that if the proposed schemes are used in real LTE deployments, there will be improvements in performances similar with what are reflected in the simulation results. Thus, the materialization of the proposed methods seems warranted.

Chapter 1

Introduction

Cellular communication has become an essential part of our daily life. In the last decades, it has evolved from being an expensive technology for a few selected individuals to today's ubiquitous systems used by a majority of the world's population. The rapidly growing demand for bandwidth in cellular communication has motivated the telecommunication standardization authorities to introduce high performing and cost effective platforms. In order to provide solutions for next generation wireless broadband, the industries have demonstrated strong support for Long Term Evolution (LTE) specified by 3GPP.

The major key performance indicators (KPIs) of the cellular services are user throughput and spectral efficiency, packet error loss rate (PELR), latency, call/session drop rate, power saving, stability in power requirement, appropriate decision for cell reselection and handover, maximum user speed, and so forth. The access strata (AS) is the lower part of the protocol stack that establishes peer-to-peer connections between the user equipment (UE) and the base station and handles functionality operating between them. The access strata (AS) oversees and performs the major functions that directly relate to the KPIs. Thus, an efficient control of AS can play a major role in the improvement of KPIs. The KPIs based on the current specifications of LTE already far outperform all other existing technologies. However, the rapidly growing demand for higher data rates and better quality of services has motivated research in AS functionalities with a view to assess and further improve the KPIs.

1.1   Basic Functionalities of Access Strata (AS)

The AS operation is broad and its better control over various activities can greatly improve the KPIs. In the control plane, the AS is comprised of the bottom three layers. Layer 1 holds the physical connection and implements modulation, coding, multiplexing, scrambling, filtering, interleaving, and so forth. Layer 2 includes three sublayers, namely, medium access control (MAC), radio link control (RLC) and packet data convergence protocol (PDCP). The major functions of layer 2 are scheduling and allocation of radio resources, segmentation, concatenation and reordering of data units, retransmission of missing data units, security of data, discontinuous reception (DRX), random access,

modulation and coding scheme (MCS) selection, synchronization, hybrid ARQ (HARQ), UE status reporting, header compression and decompression and so on. The layer 3 contains radio resource control (RRC). The layer 3 has an overall control over layer 1 and layer 2. The major functions of layer 3 are the configuration of layer 2 parameters and controlling various activities, for example, cell switching (cell reselection and handover), paging, initial attach and detach, tracking area update and transmit power level. Chapter 2 and 3 provides a limited background for understanding the functionalities of AS. Technical books depict the operation of AS in detail [1].

1.2   AS Aspects Considered for Performance Improvement

The research delves into the existing 3GPP specifications and proposes implementation techniques for access strata (AS) in LTE, analyze them, formulate necessary models and looks for the scopes of modifications with a view to achieve improvement in key performance indicators (KPIs). For this purpose, the research addresses a few specific aspects of the access strata (AS). This includes serving and neighbor cell measurement by physical layer, power saving operation by MAC layer, resource allocation by MAC layer, packet retransmissions for error correction by RLC layer, RRC control over cell switching and RRC configurations for layer 1 and layer 2.

1.3   Background and Motivation

The research presented in this dissertation aims at improvement in KPIs for LTE in various ways. It envisions improvement in power saving of the UE when delay tolerant applications are active and less mean delay in service resumption after a long inactivity of the user. The research identifies the tradeoff between power saving and packet delay for discontinuous reception operated by MAC layer. It finds out the overall delay in radio link control (RLC) retransmissions for nonreal-time (NRT) traffic. It then estimates how much relax in the delay requirement is permissible.

Some research has already investigated the impact of parameter values on power saving and delay for discontinuous reception. It is shown that a good power saving can be achieved at the cost of delay. Therefore, some research proposes algorithms to find proper parameter values for maximizing power saving while satisfying a specified delay constraint as well as minimizing delay while satisfying a specified power saving. Some research

proposes an algorithm to select appropriate configuration while keeping the delay below a threshold. Some research also shows that continuous adaptation of parameters can obtain better performance compared to fixed parameters. Thus, adaptation of parameters based on the current traffic is suggested.

The research presented in this dissertation proposes techniques for power saving that avoid the continuous estimation of traffic [2]. In fact, the estimation of current traffic, used in many proposals as mentioned earlier, increases complicacy and burden. The proposal exploits the permissible relax in the delay requirement to improve power saving by MAC layer operation. It also attempts to reduce delay in service resumption after a long inactivity. The research performs stochastic analysis in terms of power saving and packet delay using packet arrivals according to Poisson distribution.

The RRC layer in the AS is responsible for the decisions with regard to cell switching in both idle state and when a call or session is up. In this process, the layer 1 of the UE performs downlink signal measurements and sends measurement reports to layer 3 of the UE and the base station. In a network with base stations of different capacities, a high number of cell switching needs to be performed to support seamless mobility, and there are difficulties in setting proper timing for the triggering of cell switching when the user moves at a very high speed. A scaling of the cell switching parameters is used depending on the user speed. The proper timing depends on the scenario and the scenarios may vary widely as a result of differences in user speed, cell shapes, cell sizes, position and traffic pattern of highways and other streets with respect to the cellular coverage, operating frequency, multipath environment and fading condition, and so forth.

Some research has already optimized the cell switching parameters to achieve robust and seamless mobility of UEs in a network with base stations of different capacities. The existing research performs mobility state estimation to determine the velocity of the UEs. It uses context-aware mobility management for high speed UEs. Some research shows that the adjustment of cell switching parameters with higher level of granularity improves the performance. There are many proposals related to the adjustment of cell switching parameters so far. But there is still deficiency in the controllability that is required for speedy users.

The research presented in this dissertation proposes techniques for adjustment of cell switching parameters. Instead of letting eNodeB gain the whole control over adjustment, a partial control is established at the UE, which can be accomplished quickly using no overhead. The UE controlled adjustment can function as fine-tuning of the scaling of cell switching parameters. The research also incorporates Doppler spread based velocity

estimation in order to ensure the current state of the user and mitigate the chances of unwanted effects. The research employs a versatile controllability over cell switching to address various challenges. It formulates the proposed model of scaling cell switching parameters using log-distance path loss model.

The research presented in this dissertation also identifies various interferences in a network with base stations of different capacities [3]. It explores interference mitigation techniques using fractional frequency reuse schemes and checks the scopes of improvement in spectral efficiency. It derives relevant system model and formulates the problem for maximum throughput while looking for optimum sub-carrier allocation to the users in each of the sectors and layers in the cell. It derives possible arrangements of sectors and layers for gradual increase and decrease in resource allocation by MAC layer with distance from the base station. It suggests useful signaling by RRC layer for coordination. It derives possible techniques for the distribution of resource share among sectors and layers to better mitigate the interferences. Based on these techniques, it finds out suitable sectorization and layering and develops a near-optimal algorithm for resource distribution among sectors and layers.

The research develops MATLAB codes for the purpose of validation of benefits and performance improvements of the proposed methods. It benchmarks the proposed schemes against the existing schemes based on the simulation results. The simulation results manifest significant improvement in KPIs in the proposed schemes.

All cellular operators around the world, except WiMAX operators, have now LTE as their target in the technology road map. Moreover, many WiMAX operators are pondering switching to LTE. Thus, it is widely anticipated that there will be billions of LTE users soon. These users are going to receive LTE services for a good part of their 24 hours and considering the emerging applications, the LTE services will be used for many different valuable purposes. Thus, it acted as a source of motivation for the research in LTE technologies that a small contribution to the improvement in LTE technologies has a direct bearing on the innumerable moments of billions of lives. Initially, performance analyses of a few different aspects of AS of LTE were performed in the research, which demonstrated AS as overpowering [4]−[6]. This further motivated a major research in this area.

Since LTE is a new technology, a discussion on the basic concepts of LTE should be incorporated and so, it is provided in the next chapter. The discussion focuses on AS related aspects.

Chapter 2

Overview of LTE

GSM and its evolution through GPRS, EDGE, WCDMA and HSPA, is the technology stream of choice for the vast majority of the world's mobile operators. In order to meet the continued traffic growth demands, an extensive effort has been made by the 3G Partnership Project (3GPP) to develop a new standard for the evolution of GSM/HSPA technology towards a packet optimized system, referred to as Long Term Evolution (LTE). The later version of LTE is known as LTE-Advanced. This chapter provides a brief overview of basic LTE technologies as well as a few LTE-Advanced features. A detailed discussion is available in [1].

2.1   Introduction

The cellular wireless communications industry witnessed tremendous growth in the past decade with over four billion wireless subscribers worldwide. But at present, LTE is considered the most promising cellular technology and it is expected to meet the growing demand of data rate. The massive deployment of LTE is already under way around the world. By the end of 2015, more than 440 operators had commercially launched LTE coverage in more than 150 countries. 97 LTE-Advanced systems were launched in 48 different countries, while more than 140 operators are investing in LTE-Advanced in more than 60 countries [7].

The first generation (1G) analog cellular systems supported voice communication with limited roaming. The second generation (2G) digital systems promised higher capacity and better voice quality than did their analog counterparts. The two widely deployed second-generation (2G) cellular systems are GSM (global system for mobile communications) and CDMA (code division multiple access). The early 2G systems were primarily designed to support voice communication. In later releases of these standards, capabilities were introduced to support data transmission. However, the data rates were generally lower than that supported by dial-up connections. The ITU-R initiative on IMT-2000 (international mobile telecommunications 2000) paved the way for evolution to 3G. Aset of requirements such as a peak data rate of 2 Mbps and support for vehicular mobility were published under IMT-2000 initiative. Both the GSM and CDMA camps formed their own separate 3G

partnership projects (3GPP and 3GPP2, respectively) to develop IMT-2000 compliant standards based on the CDMA technology. The 3G standard in 3GPP is referred to as wideband CDMA (WCDMA) because it uses a larger 5 MHz bandwidth relative to 1.25 MHz bandwidth used in 3GPP2's cdma2000 system.

Table 2.1 Evolution of 3GPP technologies

Generation	Technologies	Service	Theoretical Data Rate	Multiplexing Technique
2G	GSM	Voice	9.6-19.2 kbps	FDMA/TDMA
2.5G	GPRS	Data	44-171.2 kbps	
2.75G	EDGE	Data	384 kbps	
3G	UMTS (WCDMA)	Voice /Data	144 -1920 kbps	CDMA
3.5G	HSDPA	Data	14 Mbps (DL)	
	HSUPA	Data	5.76 Mbps (UL)	
3.75G	HSPA Evolution or HSPA+	Data	42 Mbps (DL) 12 Mbps (UL)	
3.9G	LTE	Voice /Data	100 Mbps (DL) 50 Mbps (UL)	OFDMA
4G	LTE-Advanced	Voice /Data	1 Gbps (DL) 500 Mbps (UL)	

The first release of the 3G standards did not fulfill its promise of high-speed data transmissions. A serious effort was then made to enhance the 3G systems for efficient data support. The 3GPP2 first introduced the CDMA 2000 1x EV-DO system. The 3GPP followed a similar path and introduced HSPA (high speed packet access), which is an enhancement to the WCDMA system. The HSPA standard reused many of the same data-optimized techniques as the EV-DO system.

Table 2.2 Evolution of 3GPP2 technologies

Generation	Technologies	Service	Theoretical Data Rate	Multiplexing Technique
2G	CDMAOne (IS-95 A)	Voice	14.4 kbps	CDMA
2.5G	CDMAOne (IS-95 B)	Voice	64 kbps	
2.75G/3G	CDMA 2000 1x RTT [Rel 0, Rel A, Rel B, 1x Rel C]	Voice /Data	153.6 kbps	
3G	CDMA 2000 1x EV-DO Rev 0	Data	2.4 Mbps (DL) 153.6 kbps (UL)	
3.5G/3.75G	CDMA 2000 1x EV-DO Rev A	Data	3.1 Mbps (DL) 1.8 Mbps (UL)	
	CDMA 2000 1x EV-DO Rev B	Data	9.3 Mbps (DL) 5.4 Mbps (UL)	
3.9G/4G	CDMA 2000 1x EV-DO Rev C or UMB was first adopted but later, LTE/ LTE-Advanced has been adopted instead.	Data	100 Mbps (DL) 50 Mbps (UL) or 1 Gbps (DL) 500 Mbps (UL)	OFDMA

While HSPA and EV-DO systems were being developed and deployed, IEEE introduced the IEEE 802.16e standard for mobile broadband wireless access. This standard was introduced as an enhancement to an earlier IEEE 802.16 standard for fixed broadband wireless access. The 802.16e standard employed a different access technology named orthogonal frequency division multiple access (OFDMA) and claimed better data rates and spectral efficiency than that provided by HSPA and EV-DO. Although the IEEE 802.16 family of standards is officially called Wireless MAN, it has been dubbed WiMAX (worldwide interoperability for microwave access) by an industry group named the WiMAX Forum. The WiMAX system supporting mobility as in IEEE 802.16e standard is referred to as Mobile WiMAX. The introduction of Mobile WiMAX led both 3GPP and 3GPP2 to develop their own version of beyond 3G systems. The Long Term Evolution (LTE) refers to the 3GPP Evolved UMTS Terrestrial Radio Access technology or E-UTRA and this is the

beyond 3G system in 3GPP. Its first version was documented in Release 8 of 3GPP specifications. Release 9 enhanced specifications for various features and also, introduced several new features. The successor to LTE is LTE-Advanced (LTE-A). LTE-A was first standardized in Release 10. Later, Release 11 and 12 further standardized LTE-A and LTE-A is currently being standardized in Release 13. The advent of LTE is regarded as the latest upgrade in 3GPP technologies after GSM/GPRS/EDGE and UMTS/HSPA/HSPA+ as demonstrated in Table 2.1. Here, the whole terms for the technologies are Global System for Mobile Communications (GSM), General Packet Radio Services (GPRS), Enhanced Data rates for GSM Evolution (EDGE) and Universal Mobile Telecommunications System (UMTS).

As demonstrated in Table 2.2, 3GPP2 first chose Ultra Mobile Broadband (UMB) as 3.9G/4G technologies after IS-95, CDMA 2000 and EV-DO Rev 0/Rev A/Rev B. But later, 3GPP2 stopped the development of UMB and adopted LTE/ LTE-Advanced instead.

The goal of the LTE standard is to create specifications for a new radio-access technology geared to higher data rates, low latency and greater spectral efficiency. LTE is designed to achieve peak data rate of 100 Mbps in downlink and 50 Mbps in uplink for 20 MHz bandwidth operation. However, the actual data rate is affected by radio link condition, the number of users in the cell and various overhead. The overhead includes guard band of the channel bandwidth, cyclic prefix, time windowing and the control signaling at different layers of protocol stack. LTE can achieve peak spectral efficiency 5 bps/Hz in downlink and 2.5 bps/Hz in uplink using code rates exceeding 0.5. LTE uses optimized headers and control signaling in different layers with a view to attain small difference between physical layer and application layer throughput. LTE supports QPSK, 16-QAM, and 64-QAM for data in both downlink and uplink (HSDPA supports QPSK and 16-QAM, HSUPA supports BPSK, and HSPA+ additionally supports 64-QAM in downlink and 16-QAM in uplink). LTE uses all-IP network and it supports advanced multiple antenna technologies. LTE uses optimized signaling and it offers enhanced support for end to end QoS. It also offers reduced complexity in UE and network and thus, allows reduced CAPEX and OPEX. The user plane latency is expected to be less than 5 ms for small IP packets. The user plane latency is critical, especially, for interactive applications, e.g. VoIP and gaming. LTE is optimized for user speed up to 15 km/h. But LTE supports high performance for user speed up to 120 km/h and it sustains mobility for user speed up to 350 km/h or even up to 500 km/h. The coverage is expected to be very good up to 5 km. The coverage can be good enough up to 30 km and it may range anyway up to 100 km. However, LTE actually has coverage issue because of its

high data rate support. This makes the maximum allowable path loss (MAPL) for LTE less than other 3GPP/3GPP2 technologies.

LTE-Advanced meets or surpasses the requirements of IMT-Advanced, which is regarded as 4G. LTE-Advanced has the target peak data rate 1 Gbps in downlink and 500 Mbps in uplink. The target peak spectral efficiency is 30 bps/Hz in downlink and 15 bps/Hz in uplink. LTE-Advanced will be backward compatible with LTE (i.e. LTE based terminal and network will be operable with LTE-Advanced based network and terminal, respectively).

2.2   Network Architecture

The Evolved Packet System (EPS) refers to the evolution of 3GPP UMTS radio access, packet core and its integration to legacy 3GPP and non-3GPP network, i.e. EPS refers to what is termed as LTE network in common speech. EPS provides a data flow path with particular Quality of Service (QoS). The information transferred within EPS has two major types as follows.

1. Control Information
2. User Data

Fig. 2.1 shows the network architecture for EPS [1]. EPS includes the following entities.

1. Radio Access Network known as Evolved UMTS Terrestrial Radio Access Network (E-UTRAN).
2. The non-radio aspects known as System Architecture Evolution (SAE). This includes the core network called Evolved Packet Core (EPC).

2.2.1   *Evolved UMTS Terrestrial Radio Access Network (E-UTRAN)*

The Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) includes only the base station called eNodeB. Unlike UMTS, there is no central node like RNC and a flat hierarchy is rather used. Therefore, it is said to have a flat architecture. The flat architecture creates fewer nodes in network and causes lower latency. The primary functions of the eNodeB are as follows.

− Radio bearer control and radio admission control
− Connection mobility control

- Uplink and downlink radio resource scheduling and allocation
- IP header compression and encryption of user data
- Selection of an MME during UE attachment if it is needed.
- Routing of user plane data towards Serving Gateway
- Scheduling and transmission of paging messages
- Scheduling and transmission of broadcast information and Public Warning System (PWS) message
- Measurement and reporting configuration
- Handling Closed Subscriber Group (CSG)
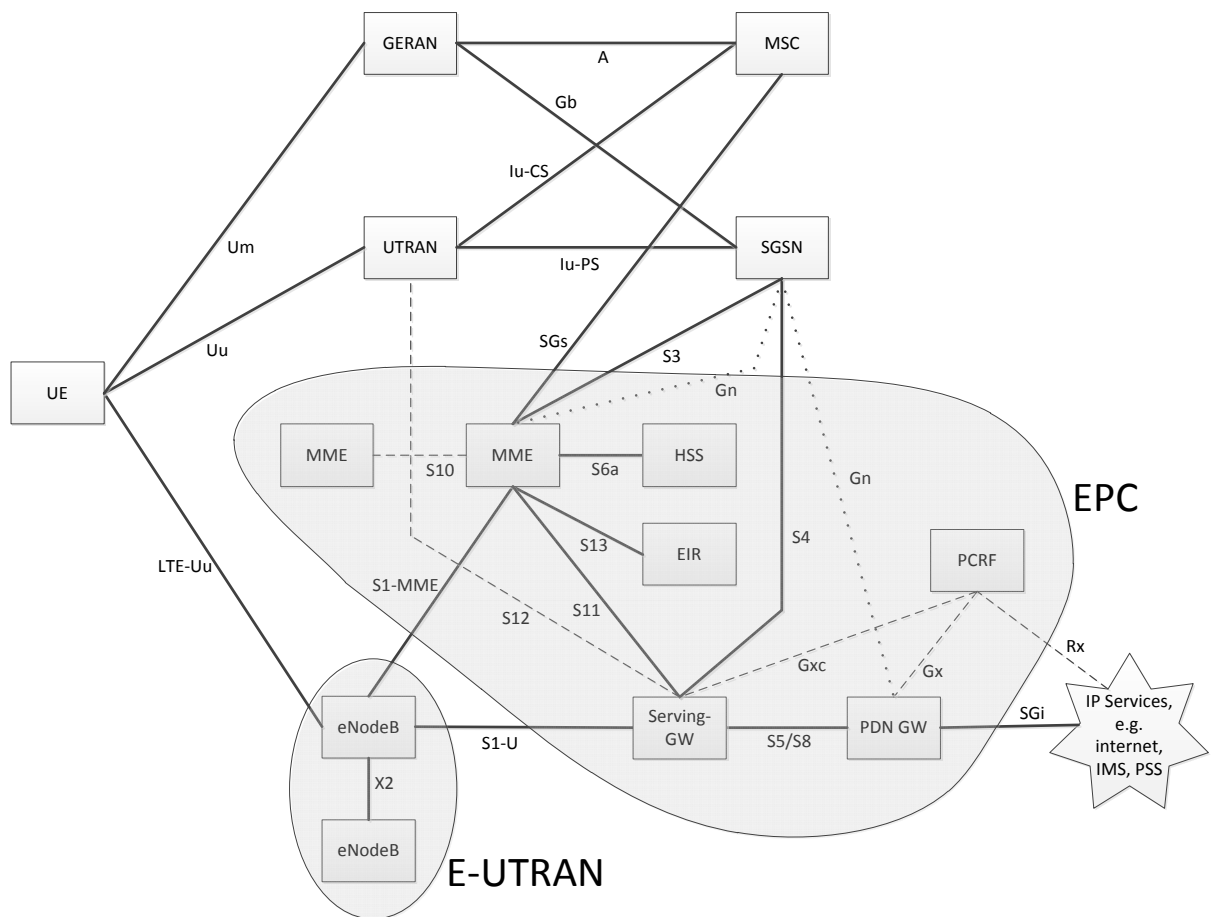
Figure 2.1: Network architecture for EPS

<u>Interfaces with eNodeB</u>

1. *S1 Interface:* The eNodeB is connected with EPC via S1 interface. An eNodeB can be connected with more than one MME/S-GW over multiple S1 interfaces. The S1 interface is split into two interfaces as shown below.

i. S1-U: S1-U is the part of S1 interface for the user plane. S1-U connects the eNodeB with the Serving Gateway (S-GW). Separate GTP-U tunnels are established for each EPS bearer for each UE.

ii. S1-MME: S1-MME is the part of S1 interface for the control plane. S1-MME connects the eNodeB with the MME. As shown in Figure 1.5, the S1-AP protocol runs over SCTP/IP on S1 interface.

2. *X2 Interface:* The neighboring eNodeBs are interconnected via X2 interfaces. The X2 interface is used for handover, load balancing, Inter-Cell Interference Coordination (ICIC), etc. The X2 interface is split into two interfaces as shown below.

i. X2-U: X2-U is the part of X2 interface for the user plane.

ii. X2-CP: X2-CP is the part of X2 interface for the control plane.

3. *LTE-Uu Interface:* The eNodeB communicates with the UE over an air interface known as LTE-Uu interface.

2.2.2   *Evolved Packet Core (EPC)*

The Evolved Packet Core (EPC) includes the following major components.

1. *PDN Gateway (P-GW):* The PDN Gateway is the gateway to interface the network towards the Packet Data Network (PDN) and thus, provides connectivity to the IP backbone. The association between the UE and the PDN is known as PDN Connection. One IP address is used for a particular PDN connection. The primary functions of the PDN GW are as follows.

   − IP address allocation for UEs.

   − Enforcement of data rate of the UE based on APN-AMBR for non-GBR bearers and based on MBR for GBR bearer.

   − Downlink packet filtering based on downlink TFT.

   − Uplink and downlink service level charging

   − Serving as mobility anchor during handover with CDMA2000 network.

2. *Serving Gateway (S-GW):* The primary functions of the Serving Gateway are as follows.

   − Routing and forwarding of IP packets.

   − Downlink packet buffering.

   − Serving as local mobility anchor during handover between eNodeBs or handover with GSM or UMTS network.

- Uplink and downlink charging per UE, PDN and QCI

- Initiation of network triggered service request procedure.

3. *Mobility Management Entity (MME):* The primary functions of the MME are as follows.

- NAS layer signaling

- NAS and AS layer security and authentication

- Management of Tracking Area List

- Selection of PDN GW and Serving GW

- Selection of MME for handover with MME change

- Selection of Serving GPRS Support Node (SGSN) selection for handover to GSM or UMTS network

- Supporting Public Warning System (PWS) message transmission

- Roaming

- EPS bearer management

- CS fallback functions

4. *Home Subscription Server (HSS):* The Home Subscription Server (HSS) stores the subscription data for the subscribers of an operator. It stores a long term key used for authentication. The HSS is updated with the information of currently serving MME of the UE. It also stores maximum bit rate restrictions for the UE and the APN, QoS information for different EPS bearers and the Visitor PLMNs in which roaming is allowed. The HSS typically includes an Authentication Centre (AuC) which authenticates the UEs. The HSS is located in the home network of the subscriber and is connected to the MME via S6a interface. The HSS uses Diameter protocol at S6a interface instead of SS7. The Diameter protocol supports transfer of subscription and authentication information in order to authorize the access of a user.

5. *Policy Control and Charging Rules Function (PCRF):* The Policy Control Enforcement Function (PCEF) is located in P-GW. Each service data flow corresponds to a defined IP packet flow through the PCEF. Policy and Charging Control (PCC) rule is set up for each service data flow depending on the subscription of the user. The PCC rule consists of a set of information to identify the service data flow and provide parameters for policy control and charging control of the service data flow. PCC includes the following.

- Packet flow based charging: This performs charging for each service data flow
- Policy control: This includes QoS control, QoS signaling, etc.

Policy Control and Charging Rules Function (PCRF) is used to provide PCEF with the PCC rule for policy and charging control. Thus, PCRF makes PCC related decisions.

6. *Equipment Identity Register (EIR):* The Equipment Identity Register (EIR) keeps record about UEs in the network and may maintain the following three lists of IMEIs.

   i. White List: The white list contains the list of IMEIs of UEs that are permitted for use.

   ii. Black List: The black list contains the list of IMEIs of UEs that are permitted for use and so need to be barred.

   iii. Grey List: There may be a grey list which contains the list of IMEIs of UEs that are permitted for use but need to be tracked by the network for evaluation or other purposes.

Interfaces in EPC

1. *S5/S8 Interface:* The Serving Gateway (S-GW) is connected with PDN Gateway (P-GW) via S5 interface when the UE is not roaming and via S8 interface when the UE is roaming in Visitor PLMN (VPLMN). The S5/S8 interface uses GPRS Tunneling Protocol (GTP) or Proxy Mobile IP (PMIP). If GTP is used, then one GTP-C tunnel is established per PDN connection for each UE. Also, separate GTP-U tunnels are established for each EPS bearer for each UE.

2. *S6a Interface:* It connects MME with HSS. It uses Diameter protocol. It uses SCTP or TCP. It supports subscription and security control between MME and HSS. The MME uses this interface to indicate the UE location. Then the HSS uses this interface to send subscription and security data for the UE.

3. *S10 Interface:* It inter-connects two MMEs. It supports mobility between MMEs. It uses GTP-C tunnel.

4. *S11 Interface:* It connects MME with S-GW. One GTP-C tunnel is established per PDN connection for each UE. It supports mobility and EPS bearer management.

5. *S13 Interface:* It connects MME with EIR.

6. *Gx Interface:* It connects P-GW with PCRF.

7. *Gxc Interface:* It optionally exists to connect S-GW with PCRF.

8. *SGi Interface:* It connects P-GW with external packet data network. Alternatively, it can connect P-GW with an internal packet data network, for example, for IMS services.

9. *SGs Interface:* It connects MME with MSC. It uses Gs protocols and supports mobility management and paging procedures between EPS and CS domain.

10. *S3 Interface:* It connects MME with Serving GPRS Support Node (SGSN). It carries control information using GTP-C tunnel to support switching between LTE and UMTS technologies and Circuit Switched Fallback (CSFB).

11. *S4 Interface:* It connects S-GW with SGSN. It carries control information using GTP-C tunnel and it carries user data using GTP-U tunnel.

12. *S12 Interface:* The S12 interface can optionally exist between S-GW and Radio Network Controller (RNC) to support direct tunneling. It carries user data using GTP-U tunnel. When S12 interface exists, S4 interface carries only control information and thus, S-GW has no connection with SGSN on user plane.

13. *Gn Interface:* Instead of S3, S4 and S12 interfaces, alternatively, Gn interface can be used to connect EPC with SGSN. In this case, both MME and P-GW are connected to SGSN using Gn interfaces. Thus, the legacy 3GPP interfaces are used. From UTRAN or GERAN, the P-GW acts like a Gateway GPRS Support Node (GGSN) and the MME acts like a SGSN.

2.3   Protocol Stack

The protocol stack has the following two types.

1. Control plane for transfer of control information
2. User plane for transfer of user data

The control plane performs signaling. Fig. 2.2 shows the control plane protocol stack between the UE and the network [1]. It includes layers as shown below.

1. *Access Strata (AS):* The Access Strata (AS) includes the following layers. The AS protocols run between the UE and the eNodeB. The operation of AS will be briefly explained in Chapter 3.

    I.   Layer 3: This layer includes only the Radio Resource Control (RRC) layer.

14

II. Layer 2: This layer includes the following sub-layers.

      i.   Packet Data Convergence Protocol (PDCP)

      ii.   Radio Link Control (RLC)

      iii.   Medium Access Control (MAC)

III. Layer 1: This layer includes only the physical layer.
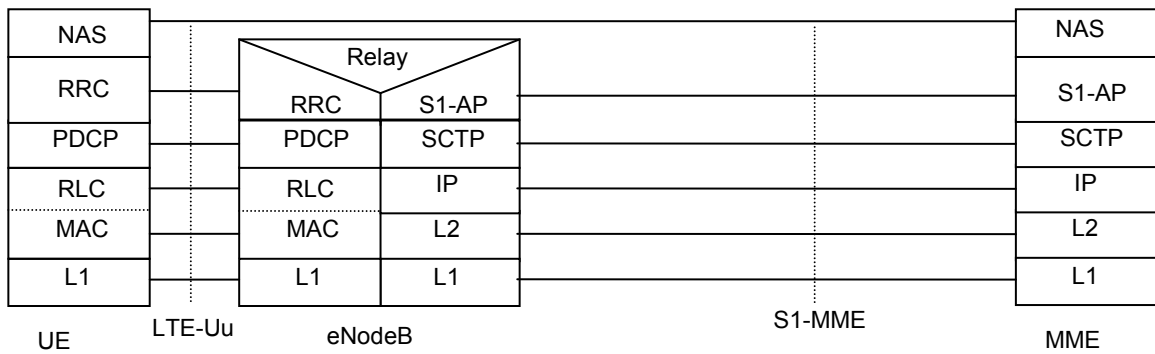
Figure 2.2: Control plane protocol stack between UE and EPC via E-UTRAN

2. *Non-Access Stratum (NAS):* The Non-Access Stratum (NAS) layer is located above the AS layer. The NAS protocols run between the UE and the MME. A NAS signaling connection is defined between the UE and the MME. The NAS layer includes the following major protocols.

    I.   EPS Mobility Management (EMM): The EMM protocols are used for attach, detach and tracking area update procedure. It is also used for UE identification, UE authentication and security mode control. In addition, it supports EMM connection related services, e.g. service request, paging.

    II.   EPS Session Management (ESM): The ESM protocols are used for activation, deactivation and modification of default or dedicated EPS bearers.

The user plane conveys the actual data traffic. The Access Strata (AS) in the user plane consists of the following layers.

    I.   **Layer 2:** This layer includes the following sub-layers.

      i.   Packet Data Convergence Protocol (PDCP)

      ii.   Radio Link Control (RLC)

      iii.   Medium Access Control (MAC)

II.   **Layer 1:** This layer includes only the physical layer.

Fig. 2.3 shows the user plane protocol stack between the UE and the EPC [1].
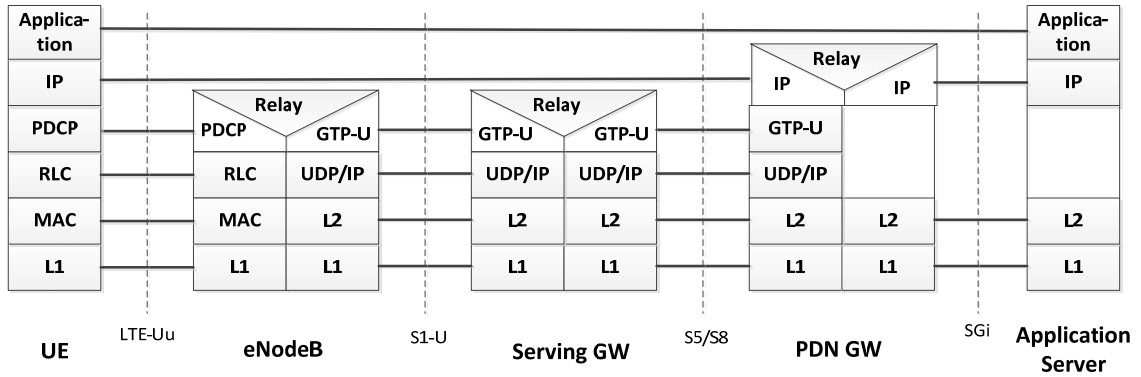
Figure 2.3: User plane protocol stack between UE and EPC via E-UTRAN

2.4   EPS Bearer

An EPS bearer is established between the UE and the PDN GW to carry user data and the EPS bearer carries traffic flows that receive a common QoS treatment between the UE and the PDN GW. All traffic mapped on the same EPS bearer receive the same treatment in terms of scheduling policy, queue management policy, rate shaping policy, RLC configuration, etc. Providing different bearer level packet forwarding treatment requires separate EPS bearers.

A default EPS bearer is established between the UE and the PDN GW and this enables always-on IP connectivity to the UE. The default EPS bearer remains established throughout the lifetime of the PDN connection. A UE can have 8 data EPS bearers established at the maximum per PDN. Besides, there can be 3 signaling EPS bearers established.

The EPS bearer can be categorized as GBR bearer and non-GBR bearer based on QoS profile. An EPS bearer is a GBR bearer if it is expected to provide a particular bit rate known as Guaranteed Bit Rate (GBR). An EPS bearer with no reserved or guaranteed bit rate resources is known as a non-GBR bearer.

The EPS bearer QoS profile characterizes an EPS bearer and it includes four parameters, namely, QoS Class Identifier (QCI), Maximum Bit Rate (MBR), Guaranteed Bit Rate (GBR) and Allocation and Retention Priority (ARP). In order to notify the EPS bearer

QoS profile, the MME sends EPS Quality of Service IE to the UE when it establishes or modifies an EPS bearer. The EPS Quality of Service IE includes QCI, GBR and MBR out of the parameters of EPS bearer QoS profile. QCI, GBR and MBR can impact packet forwarding treatment, such as, scheduling and rate control. ARP defines the relative importance of a resource request. Network uses this priority level to decide if a bearer establishment or modification request can be accepted when there is limited resource. ARP is not sent to the UE and it is rather used within the network.

Table 2.3 Characteristics for different QCI values

GBR/ Non-GBR	QCI	Priority	Packet Delay Budget (PDB)	Packet Error Loss Rate (PELR)	Application
GBR	1	2	100 ms	10^{-2}	Voice
	2	4	150 ms	10^{-3}	Live Streaming Video
	3	3	50 ms	10^{-3}	Real Time Gaming
	4	5	300 ms	10^{-6}	Buffered Streaming Video
Non-GBR	5	1	100 ms	10^{-6}	IMS Signaling
	6	6	300 ms	10^{-6}	Buffered Streaming Video, Progressive Video, HTTP, E-mail, Chat, FTP, P2P File Sharing, etc.
	7	7	100 ms	10^{-3}	Voice, Live Streaming Video, Interactive Gaming
	8 (Privileged subscribers)	8	300 ms	10^{-6}	Buffered Streaming Video, Progressive Video, HTTP, E-mail, Chat, FTP, P2P File Sharing, etc.
	9 (Non-Privileged subscribers)	9	300 ms	10^{-6}	Buffered Streaming Video, Progressive Video, HTTP, E-mail, Chat, FTP, P2P File Sharing, etc.

Each of the EPS bearers is associated with a particular QoS Class Identifier (QCI), which can have values from 1 through 9. Each QCI corresponds to a priority, a packet error loss rate (PELR) and a packet delay budget (PDB) as shown in Table 2.3. Here, the PDB indicates an upper limit for packet delay between the UE and the policy and charging enforcement function (PCEF) in the core network. The delay should not exceed PDB in the case of 98% of the packets that have not been dropped due to congestion.

In order to attain appropriate treatment in packet forwarding, the eNodeB can configure few settings based on the value of QCI, such as, scheduling weights, admission thresholds, queue management thresholds, link layer protocol configuration, etc. Scheduling of resources depend on the PDB. PDB can also be used for setting target HARQ operating points and other link layer configurations. The PELR is used for link layer protocol configurations. Both PDB and PELR have the same values in uplink and downlink for certain QCI.

2.5   Physical Layer Properties

The physical layer uses Orthogonal Frequency Division Multiple Access (OFDMA) for downlink transmission and Single Carrier Frequency Division Multiple Access (SC-FDMA) for uplink transmission. The OFDMA employs Orthogonal Frequency Division Multiplex (OFDM) for simultaneous access of multiple users. The OFDM breaks the available bandwidth into many narrow parts with subcarriers at the center of each of these parts as shown in Fig 2.4. Then the data symbols are modulated in parallel streams onto the subcarriers. The subcarriers are made orthogonal to each other by choosing a subcarrier spacing, $\Delta f = 1/T$ where T is the OFDM symbol period. This allows increased spectral efficiency with no Inter-Carrier Interference (ICI) among adjacent subcarriers virtually.

2.5.1   *Cyclic Prefix (CP) in OFDM Symbol*

The transmission duration for a symbol is made little longer than the symbol period by adding a Cyclic Prefix (CP) to each OFDM symbol. For this purpose, the last part of the OFDM symbol is copied and inserted at the beginning of the OFDM symbol. The duration of cyclic prefix is denoted as T_{CP} and the total transmission duration for the symbol becomes T $+T_{CP}$.
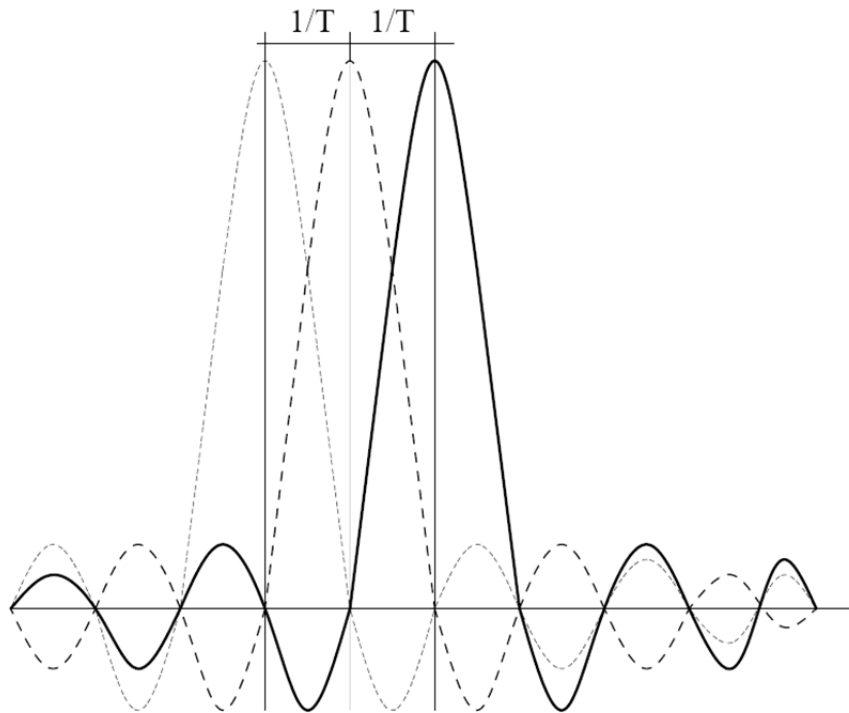
Figure 2.4: Orthogonality among subcarriers in OFDM

In cellular applications, the radio channel is time-dispersive and a symbol received from different multipath can have different delays. The delay spread can be several microseconds. Thus, a symbol received along a delayed path may overlap a subsequent symbol arriving at the receiver via a more direct path. This effect is referred to as Inter-Symbol Interference (ISI). A cyclic prefix is added to overcome ISI.

The time-dispersiveness of a radio channel is frequency dependent. Thus, a wide channel gets the different parts of its frequency range affected differently i.e. it undergoes frequency-selective fading and gives rise to distortion. A Coherence Bandwidth is considered to limit within flat fading as opposed to frequency-selective fading. The Coherence Bandwidth depends on the level of time-dispersiveness as characterized by the delay spread. OFDM uses a large number of subcarriers with narrow bandwidths instead of a wide band signal and this allows the signal on each subcarrier to be considered to undergo flat fading instead of frequency-selective fading. Thus, OFDM has the inherent property to obviate the need for an equalizer at the receiver and offer great improvement in the performance as well.

In OFDM, there can be Inter-Carrier Interference (ICI) between subcarriers in addition to Inter-Symbol Interference (ISI) due to the time-dispersiveness of the radio channel. The demodulator correlation interval for one multipath component can overlap the symbol boundary of a different multipath component. The integration interval can be chosen

for the strong multipath component but this interval may not correspond to an integer number of periods of complex exponentials for a different multipath component. This will lead to the loss of orthogonality between subcarriers and will cause ICI between subcarriers. The cyclic prefix helps overcome ICI. The advantages of the cyclic prefix can be summarized as follows.

1. The cyclic prefix allows an extension from the actual symbol period. Thus, if the span of the time dispersion of the channel is shorter than the cyclic-prefix length, a symbol received along a delayed path will not bleed into a subsequent symbol arriving at the receiver via a more direct path. Thus, cyclic prefix helps overcome ISI from the time-dispersive channel.

2. The demodulator correlation interval is kept equal to the actual symbol period and it is not increased. Assuming that the span of the time dispersion of the channel is shorter than the cyclic-prefix length, the symbol of a delayed multipath component along with its cyclic prefix will not fall into the demodulator correlation interval of the strongest multipath component. This preserves the subcarrier orthogonality at the receiver side. Thus, cyclic prefix helps overcome ICI between subcarriers.

3. The cyclic prefix allows more imperfection in the time alignment between the transmitter and the receiver. The summation of the time alignment error and the delay spread of the channel needs to be less than the cyclic prefix.

4. In case of uplink, the cyclic prefix offers an additional advantage by simplifying the implementation of the equalizer at eNodeB.

The disadvantages of the cyclic prefix can be summarized as follows.
1. The cyclic prefix increases the length of the OFDM symbol period and so, reduces the OFDM symbol rate.
2. A fraction $T/(T + T_{CP})$ of the received signal power is actually utilized by the OFDM demodulator.

2.5.2 SC-FDMA for Uplink

OFDMA offers high Peak to Average Power Ratio (PAPR). The high PAPR requires linear power amplifiers with large dynamic range. This makes the power amplifier less efficient, more expensive and more power consuming. This is a more severe problem in the

uplink compared to the downlink because it increases the cost of the UE and consumes the battery power faster. The high PAPR also degrades coverage and cell-edge performance. Therefore, in order to reduce PAPR, Single-Carrier FDMA (SC-FDMA) is used in the uplink instead of OFDMA. The time dispersive or equivalently, the frequency-selective radio channel distorts the SC-FDMA signal. Thus, an equalizer is needed to compensate for the radio channel frequency selectivity. This is the disadvantage of the use of SC-FDMA in place of OFDMA. However, the equalization is not very difficult to perform at the base station because the base station can accommodate more power and cost.

SC-FDMA signal generation is performed in the frequency domain using DFT-spread OFDM (DFTS-OFDM). DFTS-OFDM, as compared to time domain SC-FDMA generation, gives better bandwidth efficiency at the cost of little less PAPR reduction. In case of OFDMA, in every block, each subcarrier contains information about just one transmit symbol from the block. But here in case of DFT-spread OFDM, each subcarrier has information of all transmit symbols in the block as the transmit symbols were first spread by the DFT transform over the available subcarriers. Thus, if there is a deep fade on one of the subcarriers, it contains just part of a transmit symbol information on that subcarrier and the symbol may still be recovered after IDFT at the other end.

SC-FDMA allows the use of a number of subcarriers and exactly the same kind of time-frequency resource grid in the uplink as used in downlink with OFDMA in downlink. This commonality between uplink and downlink helps in implementation and resource allocation.

2.5.3 Time-Frequency Resource Grid

A constant subcarrier spacing is used regardless of the transmission bandwidth. Different sizes of transmission bandwidths are defined by assigning different number of OFDM subcarriers. The subcarrier spacing, Δf is chosen as 15 kHz. For Multimedia Broadcast Multicast Service (MBMS) dedicated cell, a reduced subcarrier spacing, $\Delta f = 7.5$ kHz can be used. Using subcarriers in the frequency axis and symbols in the time axis based on OFDM, a time-frequency resource grid is considered as shown in Fig. 2.5 and it has the same structure in both downlink and uplink. Then multiple users are allocated different parts of the time-frequency resource grid structure. Each element in the time-frequency resource grid is called a resource element and thus, a resource element represents one subcarrier and one symbol resource. Each resource element is uniquely identified by the index pair (k, l)

where k and l are the indices for subcarriers in the frequency domain and symbols in the time domain respectively.
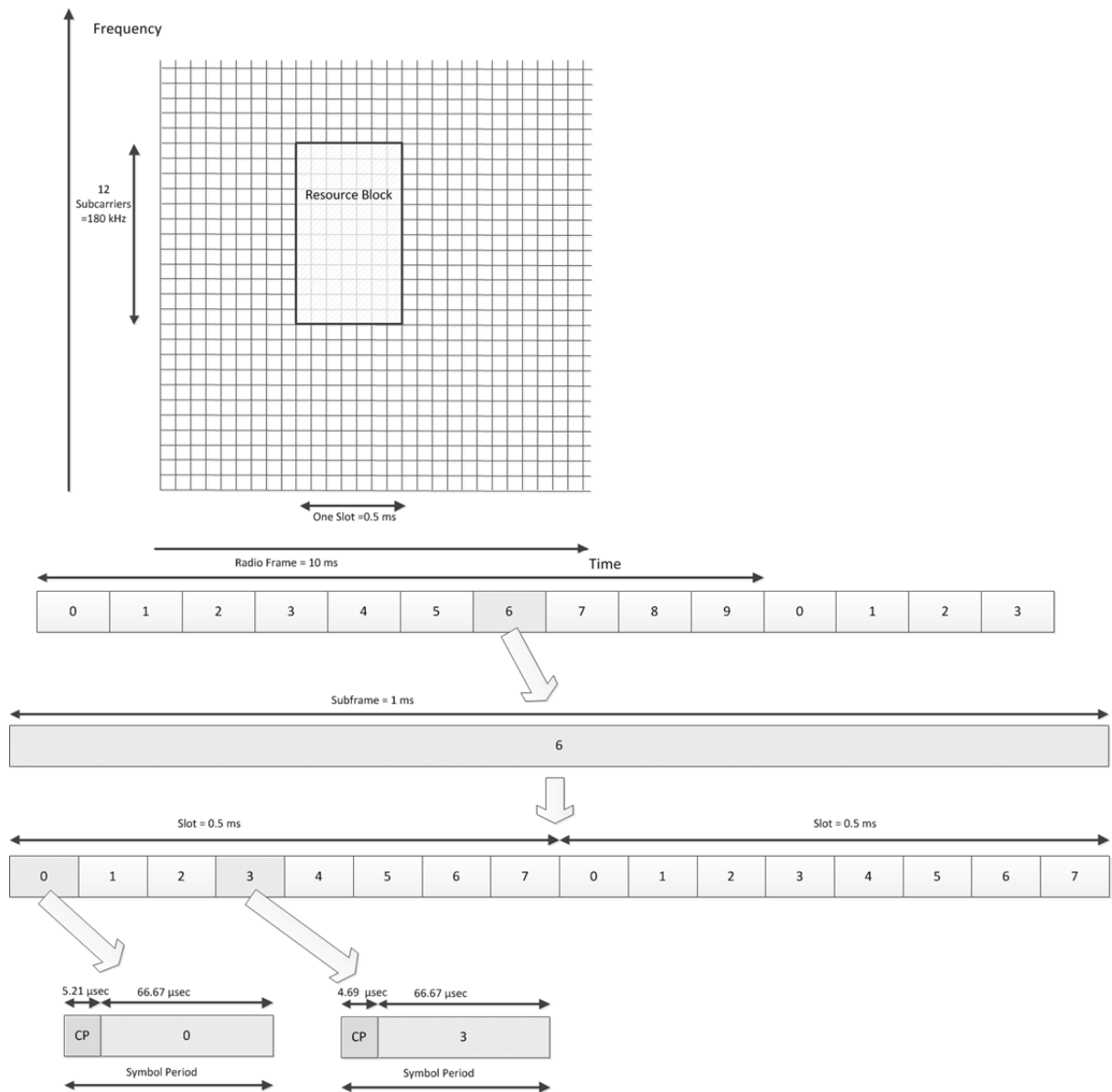
Figure 2.5: Time-frequency resource grid

To provide consistent and exact timing definitions, different time intervals within the LTE radio access specification are expressed as multiples of a basic time unit, Ts. Ts is equal to the sampling time with the largest FFT size, N = 2048 for subcarrier spacing, 15 kHz or the sampling time with FFT size, N = 4096 for subcarrier spacing, 7.5 kHz.

Ts = 1/ (2048 × Δf) = 1/ (2048 × 15000) = 1/30720000 sec ≈ 32.552 nsec.

Ts = 1/ (4096 × Δf) = 1/ (4096 × 7500) = 1/30720000 sec ≈ 32.552 nsec.

The time domain structure includes radio frames appearing one after another in time. A radio frame is defined with a length $T_{frame} = 307200 \times Ts = 10$ ms. Because of the same length, the radio frame in LTE provides easy backward compatibility with UMTS/HSPA. A radio frame consists of ten equally sized subframes. The subframes in a radio frame are numbered from 0 through 9. Each subframe is further divided into two equally sized slots.

A subframe has length $T_{subframe} = 30720 \times Ts = 1$ ms.

A slot has length $T_{slot} = 15360 \times Ts = 0.5$ ms.

A slot contains 3, 6 or 7 OFDM symbols. In uplink, there is no configuration with 3 symbols.

The radio frames appearing one after another in time are numbered sequentially and these numbers are called System Frame Number (SFN). Thus, SFN is incremented in every radio frame. SFN is 10 bits long and it is restarted after it has wrapped around. SFN helps in synchronization.

A short cyclic prefix allows for less overhead and waste of bandwidth. However, the cyclic prefix should be long enough to account for the maximum time delay of the multi-path channel. Besides, a very short cyclic prefix can cause gradual reduction in orthogonality and loss of circular convolution. The length of cyclic prefix may also take frequency reuse or interference from other cells, transmission power, etc. into account.

The following two different types of lengths are defined for cyclic prefix for 15 kHz subcarrier spacing.

1. **Normal Cyclic Prefix:** The length of normal cyclic prefix is optimized for typical urban multipath environment. This normal cyclic prefix is expected to be used most commonly. In this case, 7 OFDM symbols are used per slot.

   The length of the cyclic prefix of the 1st OFDM symbol, $T_{CP} = 160 \times Ts = 5.21$ μsec.

   The length of the cyclic prefix of the rest 6 OFDM symbols, $T_{CP} = 144 \times Ts = 4.69$ μsec.

2. **Extended Cyclic Prefix:** The length of extended cyclic prefix is optimized for larger cell size and heavy urban multipath environment. In this case, 6 OFDM symbols are used per slot.

   The length of the cyclic prefix of each OFDM symbol, $T_{CP} = 512 \times Ts = 16.67$ μsec.

The symbol period without cyclic prefix for 15 kHz subcarrier spacing is,

$T = 1/\Delta f = 1/15000 = 2048 \times Ts \approx 66.67$ μsec

The configuration for normal cyclic prefix is as follows.

The length of cyclic prefix for 1st OFDM symbol, $T_{CP} = 160 \times Ts \approx 5.21$ μsec,

The length of cyclic prefix for the rest 6 OFDM symbols, $T_{CP} = 144 \times Ts \approx 4.69$ μsec.

Thus, the total slot has length,

$T_{slot} = [(160 + 2048) + 6 \times (144 + 2048)] \times Ts = 15360 \times Ts = 0.5$ ms

The configuration for extended cyclic prefix is as follows.

The length of cyclic prefix for all OFDM symbols, $T_{CP} = 512 \times Ts \approx 16.67$ μsec,

Thus, the total slot has length, $T_{slot} = [6 \times (512 + 2048)] \times Ts = 15360 \times Ts = 0.5$ ms

A Resource Block (RB) is a basic unit for allocation of resources to the UE. A UE can be allocated integer multiple of resource blocks. In time, the length of a RB is always 1 slot which is equal to 0.5 ms. In frequency, the length of a PRB is always 180 kHz. The number of subcarriers in a resource block is denoted as N_{SC}^{RB}.

$N_{SC}^{RB} = 12$ for 15 kHz subcarrier spacing

$N_{SC}^{RB} = 24$ for 7.5 kHz subcarrier spacing

The number of available physical resource blocks in the total frequency resource is denoted as N_{RB}^{DL}. The value of N_{RB}^{DL} can be 6, 15, 25, 50, 75 or 100 depending on the channel bandwidth. The number of subcarriers in total frequency resource is, $N_{RB}^{DL}N_{SC}^{RB}$.

The number of symbols in a resource block or in a slot is denoted as N_{symb}^{DL}. Thus,

$N_{symb}^{DL} = 7$ for 15 kHz subcarrier spacing with normal cyclic prefix

$N_{symb}^{DL} = 6$ for 15 kHz subcarrier spacing with extended cyclic prefix

A RB block consists of $N_{SC}^{RB} \times N_{symb}^{DL}$ resource elements. The composition of resource elements in a resource block occur in one of the following three ways.

$12 \times 7 = 84$ resource elements for 15 kHz subcarrier spacing with normal cyclic prefix

$12 \times 6 = 72$ resource elements for 15 kHz subcarrier spacing with extended cyclic prefix.

2.5.4   Physical Channels

The physical channels at layer 1 provide air interface support for data transfer. The following downlink physical channels are available:

1. Physical Broadcast Channel (PBCH)
2. Physical Control Format Indicator Channel (PCFICH)
3. Physical Downlink Control Channel (PDCCH)
4. Physical Hybrid ARQ Indicator Channel (PHICH)
5. Physical Downlink Shared Channel (PDSCH)
6. Physical Multicast Channel (PMCH)

The following uplink physical channels are available:

1. Physical Uplink Shared Channel (PUSCH)
2. Physical Uplink Control Channel (PUCCH)
3. Physical Random Access Channel (PRACH)

2.6   Femtocell

The macro cells do not provide the most desirable performance in every scenario. For example, the indoor coverage can be very difficult in some cases. The service coverage may be limited or unavailable inside home, small business area, etc. In such cases, an access point can be set up at customer premises and it functions as a small base station as shown in Fig. 2.6 [8]. The access point is called Home eNodeB (HeNB) and such Home eNodeB cell is traditionally known as femtocell. The Home eNodeB can provide service indoors where the coverage would normally be limited or unavailable. Also, a number of Home eNodeBs can be installed in a campus or large business environment. The transmit power of Home eNodeB is typically less than 23dBm. The Home eNodeB is assigned a human readable and user friendly name. The Home eNodeB transmits the Home eNodeB name on a system information message.

The Home eNodeB is configured for operation in any of the following modes.

i. **Open Access Mode:** The Home eNodeB operates like an ordinary eNodeB. It provides services to all subscribers of any PLMN if there is proper roaming agreement.

ii.   **Closed Access Mode:** The Home eNodeB provides services only the users who belong to a Closed Subscriber Group (CSG) and it controls the access of users. The HeNB is typically operated in this closed access mode.

iii. **Hybrid Access Mode:** The Home eNodeB provides services to its associated CSG members preferentially. In addition, it provides services to subscribers of any PLMN who is not allowed to access the CSG, if there is proper roaming agreement.
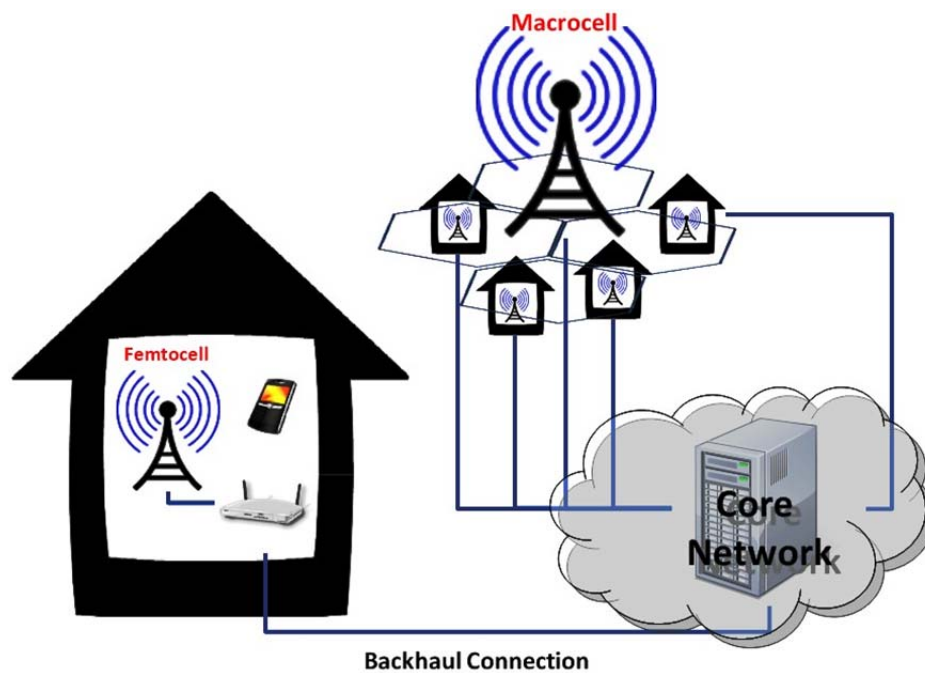
Figure 2.6: Femtocell deployment along with a macrocell

2.7   Heterogeneous Networks (HetNets)

The eNodeB can be used like a conventional base stations with typical transmit power more than 40 dBm and antenna gain between 12 and 15 dBi. The eNodeB is then called Macro eNodeB (MeNB) and the cells served are known as macro cells. A homogeneous network uses a macro cells based deployment. In this case, all eNodeBs have similar transmit power levels, antenna patterns and similar backhaul connectivity to the data network. As traffic demand grows, network relies on cell splitting or additional carriers to overcome capacity and link budget limitations and to maintain uniform user experience. This process is complex and iterative. Moreover, site acquisition for macro base-stations with towers becomes more difficult in dense urban areas. Thus, a more flexible deployment model is

needed for operators to improve broadband user experience in ubiquitous and cost effective way. A solution is introduced in Release 10 as an LTE-Advanced feature, known as heterogeneous networks (HetNets).

The HetNet uses simultaneous deployment of eNodeBs with different levels of capacities as shown in Fig. 2.7 [9]. It is considered as a very promising approach to enhance network capacity, overall performance, and to increase coverage in a cost effective way. A HetNet consists of regular macro cells transmitting typically at high power level, overlaid with low-power small cells such as pico cell, femto cell, remote radio head (RRH), and relay node (RN) etc. Among these different types of deployment, only the Home eNodeB or femto cell may restrict access to the users by using closed subscriber group. The incorporation of such small cells allows offloading traffic from macro cell and providing better network experience by connecting UEs in small cells with low transmission power. A pico cell has lower transmit power and antenna gain. Its typical transmit power is between 23 dBm and 30 dBm with antenna gain up to 5 dBi.

The initial network deployment may include only macro cells. The macro cells are usually placed in a cellular network using a prudent network plan. Thereafter, pico cells, femto cells and/or relay nodes may be added with time based on subsequent requirements. The placement of low power small cells is typically based on just a knowledge of coverage issues and traffic densities in the network. The addition of pico or femto cells can enhance coverage and capacity greatly. It offers extension of the reach of coverage with very little additional backhaul expenses. It improves the conditions in coverage holes providing higher data-rates at cell edge or in hot-spots. It improves data rate for particular users who would experience poor data rate otherwise. The low power nodes are smaller base stations with lower antenna gain compared to macro cells and so, the site acquisition for their addition can also be simpler. The HetNet has become essential for providing ubiquitous broadband user throughput.

Different types of deployment scenarios for HetNets are already available. In multicarrier deployment, small cells utilize different carrier frequency than the macro cell. This process effectively reduces inter-cell interference (ICI) but does not ensure proper spectral utilization. On the other hand, co-channel deployment is utilized by using the same carrier for both of macro cell and small cell in which the spectral efficiency is increased via spatial reuse and popular deployment approach in HetNets. Though co-channel approach ensures effective spectrum utilization but causes significant inter-cell interference (ICI)

among the macro cell and small cells. Especially, the users in edge area of low power small cells are mostly vulnerable to strong interference signals from high power macro cell.
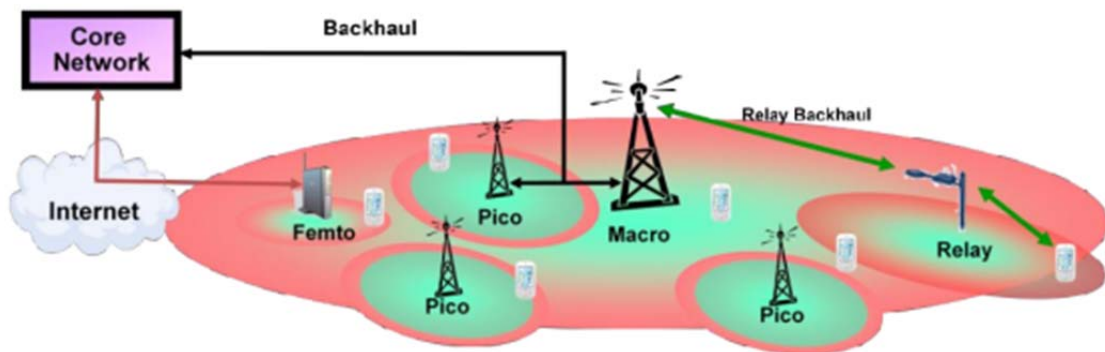
Figure 2.7: Heterogeneous Networks (HetNets)

2.7.1   Cell Range Expansion (CRE)

In HetNets, the load among tiers is unbalanced due to significant differences in transmit power levels. The UE tends to connect to the macrocell even when the path loss conditions between the small cell and the UE are better, because the handover decision is based on the received signal strength in downlink measured at a UE. As a remedy to this, the 3rd Generation Partnership Project (3GPP) standardized to virtually increase a small cell's coverage area as shown in Fig. 2.8 [10], by biasing handover decisions between the different eNBs such that UEs are handed over to pico cells earlier than usual, thus shifting load from the macro cell to the pico cell. This is known as cell range expansion (CRE). Similarly cell reselection parameters can be adapted for users in RRC_IDLE state. CRE is not only effective for optimizing the use of resources in the system, but also for reducing the frequency of handovers, hence improving system throughput and user experience. However, with CRE, the UEs handed over to the pico cell suffer from lower Signal to Interference plus Noise Ratio (SINR) than usual.
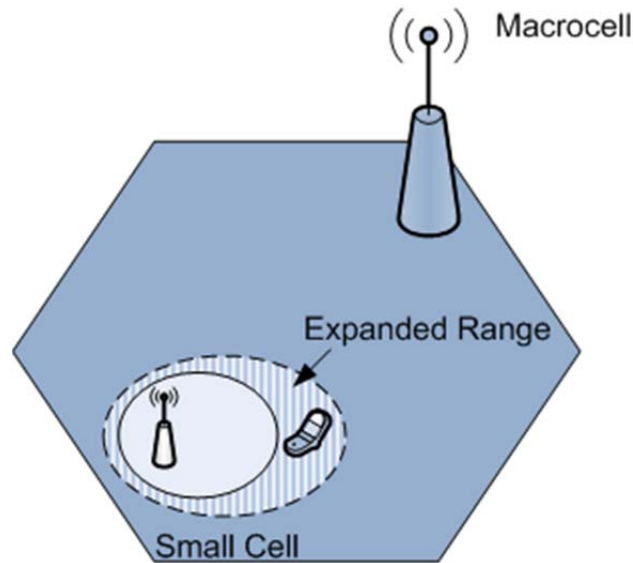
Figure 2.8: Cell Range Expansion (CRE) in HetNets

2.7.2 Interference Issues

In general, there are two types of interference in HetNets as shown below.

1. Co-tier interference: This type of interference occurs among network elements that belong to the same tier in the network. In the case of a picocell network, co-tier interference occurs between neighboring picocells.

2. Cross-tier interference: This type of interference occurs among network elements that belong to different tiers of the network, i.e. interference between picocells and macrocells.

The interference problem in HetNets is especially challenging because of the following reasons.

1. Since a large number of cell boundaries are created in HetNets and there is a large difference between the transmit power levels of macro and pico base stations, the interference problem become more challenging. In order for a UE to obtain service from a low power base station in the presence of macro base stations with stronger downlink signal strength as shown in Fig. 2.9 [11], the pico cell needs to perform interference coordination with the dominant macro interferers.

2. With CRE, the pico cells are expected to have large areas with low signal-to-interference conditions on the cell edge as shown in Section 2.7.1. This also results in a challenging RF environment for users on the cell edge.

29
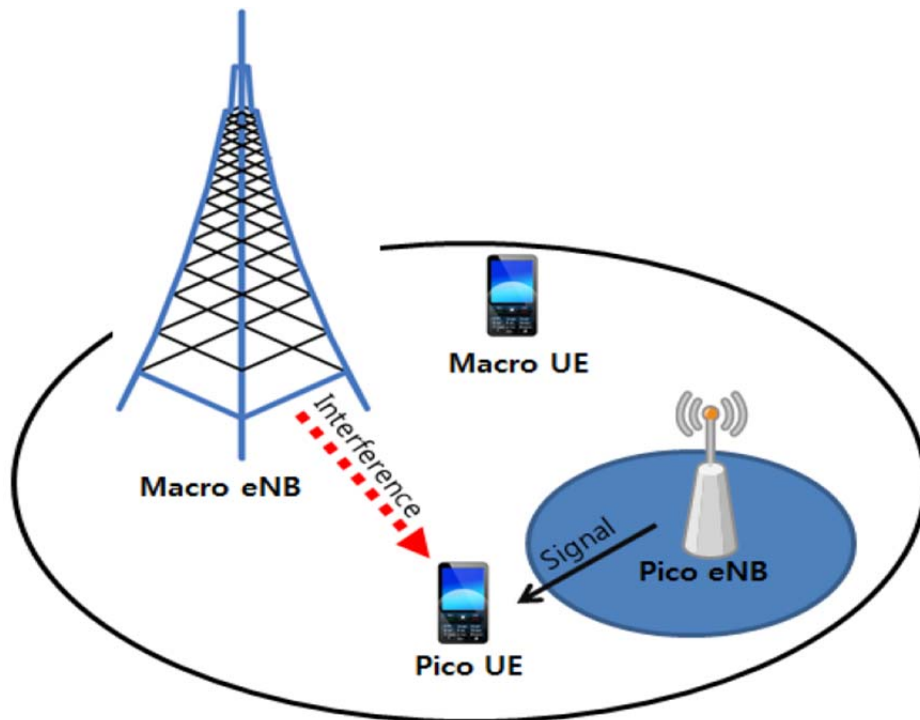
Figure 2.9: The impact of dominant macro interferer on pico UE

3. Low-power nodes, especially, femto cells are typically deployed in an ad hoc manner by users. They can even be moved or switched on/off at any time. Hence, traditional network planning and optimization becomes inefficient because operators do control either the number or the location of these cells. The unplanned deployment aggravates the interference problem.

4. Some femto cells may operate in CSG mode, in which cell access is restricted and only the owner or subscribers of the femto cells may be allowed to access the femto cells as shown in Section 2.6. For UEs that are close to these femto cells but yet barred from accessing them, the interference caused by the femto cells to the UEs can be particularly severe.

Since the interference problem is critical, Inter-Cell Interference Coordination (ICIC) plays a vital role in HetNets and various ICIC techniques are available.

2.8 Cell Switching

LTE supports mobility, i.e. the service is continued as the user moves out of the coverage of one cell and enters the coverage of another cell. The serving cell is updated automatically with the movement of the UE and thus, cell switching takes place. The

decision for the change of cell depends on the relative radio link quality between the current cell and the neighbor cells at the position of the UE. LTE uses cell reselection or handover as the method for cell switching. In the RRC_IDLE state, if the UE changes the cell on which it is camped, the cell reselection procedure is performed. On the other hand, in the RRC_CONNECTED state, if the serving cell changes, then the handover procedure is performed. The handover can also be performed to move a user from a heavily loaded cell to a lightly loaded cell for the purpose of load balancing.

The neighbor cells around the UE can be classified as follows.

1. *Intra-Frequency Cells:* The current cell and the neighbor cell operate on the same carrier frequency.

2. *Inter-Frequency Cells:* The current cell and the neighbor cell operate on a different carrier frequency.

3. *Inter-RAT Cells:* The neighbor cell operates on a Radio Access Technology (RAT) other than LTE, e.g. GERAN, UTRAN and CDMA2000.
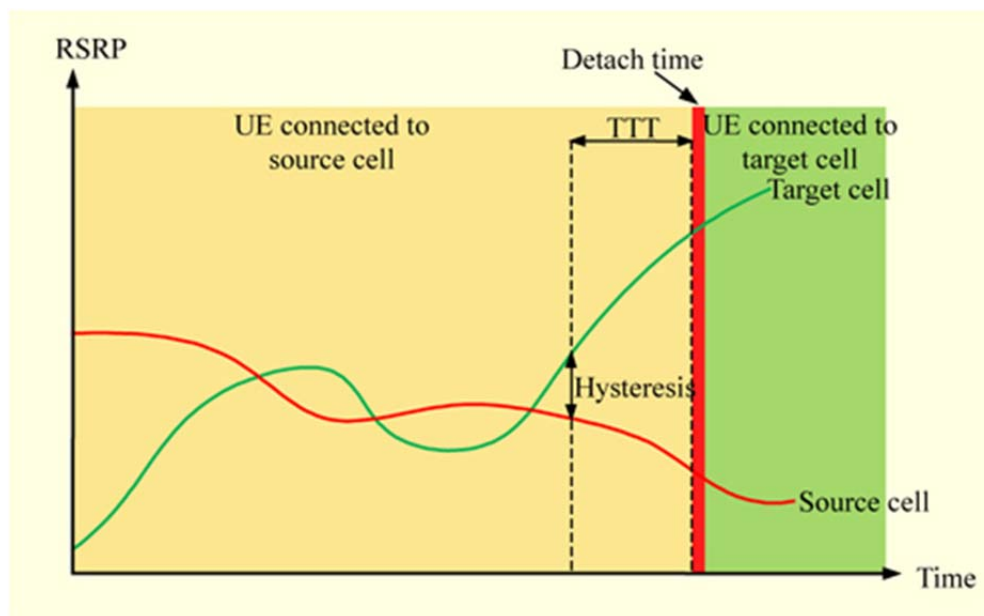
Figure 2.10: Margin in received radio link quality during handover

Because of the time varying nature of the radio signals, it is possible that what appears to be an increase or decrease of the received signal strength is actually a fast signal fluctuation that lasts for only a short period of time. Thus, the decision for cell switching must allow the target cell to remain better than the serving cell by a sufficient margin over certain period. Otherwise, there is a possibility of switching back and forth

between cells unnecessarily, which is known as ping-pong events. The unnecessary cell switching degrades the system performance using the system resources unnecessarily. The margin, shown in Fig. 2.10 [12], incorporates a hysteresis and optionally, a cell specific offset. The cell specific offset allows biasing the cell switching decision toward particular cells. The period over which the target cell remains better before cell switching is called time-to-trigger (TTT). Because of SINR fluctuation, sufficiently large values of hysteresis margin and TTT are required in order to avoid ping pong.

2.8.1 Cell Reselection

The RRC layer of the UE makes the decision for cell reselection in an attempt to operate on the best quality RF carrier. For this purpose, the layer 1 of the UE performs periodic measurements of the neighbor cells and the serving cell and reports the measurement results to the RRC layer. The RRC layer of the UE evaluates the rank of cells frequently as follows.

The rank of the serving cell, $R_s = Q_{meas,s} + Q_{hyst}$

The rank of the neighbor cell, $R_n = Q_{meas,n} - Q_{offset}$

$Q_{meas,s}$ and $Q_{meas,n}$ are the signal strength of the serving cell and the neighbor cell, respectively, measured in dBm. Q_{hyst} is the hysteresis margin. Q_{offset} is a cell specific offset. If the UE finds that a neighbor cell has been continuously better than the serving cell for $T_{reselectionRAT}$ period, the RRC layer of the UE selects the new cell for service and thus, cell reselection takes place. A separate $T_{reselectionRAT}$ timer is started for each cell that becomes better ranked than the serving cell. The UE reselects to the highest ranked cell when the corresponding timer expires. The cell reselection can be triggered if more than 1 second has elapsed since the UE camped on the current serving cell. The eNodeB configures the values of Q_{hyst}, Q_{offset} and $T_{reselectionRAT}$ using system information messages.

As the UE moves, if the cell reselection occurs too late, the radio link with the serving cell can get too poor. This can cause radio link failure (RLF). It may be noted that the above procedure applies only to intra-frequency neighbor cells and equal priority inter-frequency neighbor cells.

2.8.2 Handover

The RRC layer of the eNodeB makes the decision for handover in an attempt to operate on the best quality RF carrier. The handover is always hard handover in LTE, that is, it uses break before make approach. The UE sends measurement reports, based on which, handover decisions are made. Typically, the RRC layer of the eNodeB decides on handover after a certain number of measurement reports are received from the UE indicating the neighboring cell better than the serving cell by a sufficient hysteresis margin. The measurement reporting can be either event-triggered or periodic as shown below.

1. *Event-Triggered:* The UE is considered to enter and to leave a particular event when certain conditions are met. The UE begins sending measurement reports when the entering condition for a particular event is met for *TimeToTrigger* period As long as the event remains activated, the UE keeps on sending measurement reports one after another at a certain interval, *ReportInterval* until it reaches a maximum number of measurement reports. The UE can optionally send a measurement report when the leaving condition for an event is met. There are five types of events for measurement reporting of LTE cells. Their basic indications are shown below.

    i.   *Event A1:* The serving cell becomes better than a threshold.

    ii. *Event A2:* The serving cell becomes worse than a threshold.

    iii. *Event A3:* The neighbor cell becomes better than the serving cell by an offset.

    iv. *Event A4:* The neighbor cell becomes better than a threshold.

    v.   *Event A5:* The serving cell becomes worse than a threshold, threshold 1 and the neighbor cell becomes better than another threshold, threshold 2.

Fig 2.11 and Fig 2.12 explain event A1 and event A3 based measurement reporting for exemplification [13]. Their whole entering condition is shown below. The entering condition includes hysteresis for all five events.

Event A1: Serving cell measurement $-$ Hysteresis $>$ Threshold

Event A3:

Neighbor cell measurement $+$ OffsetFreq for neighbor cell

$\quad+$ *CellIndividualOffset for neighbor cell $-$ Hysteresis*

$\quad>$ *Serving cell measurement $+$ OffsetFreq for serving cell*

$\quad+$ *CellIndividualOffset for serving cell $+ a3 - Offset*
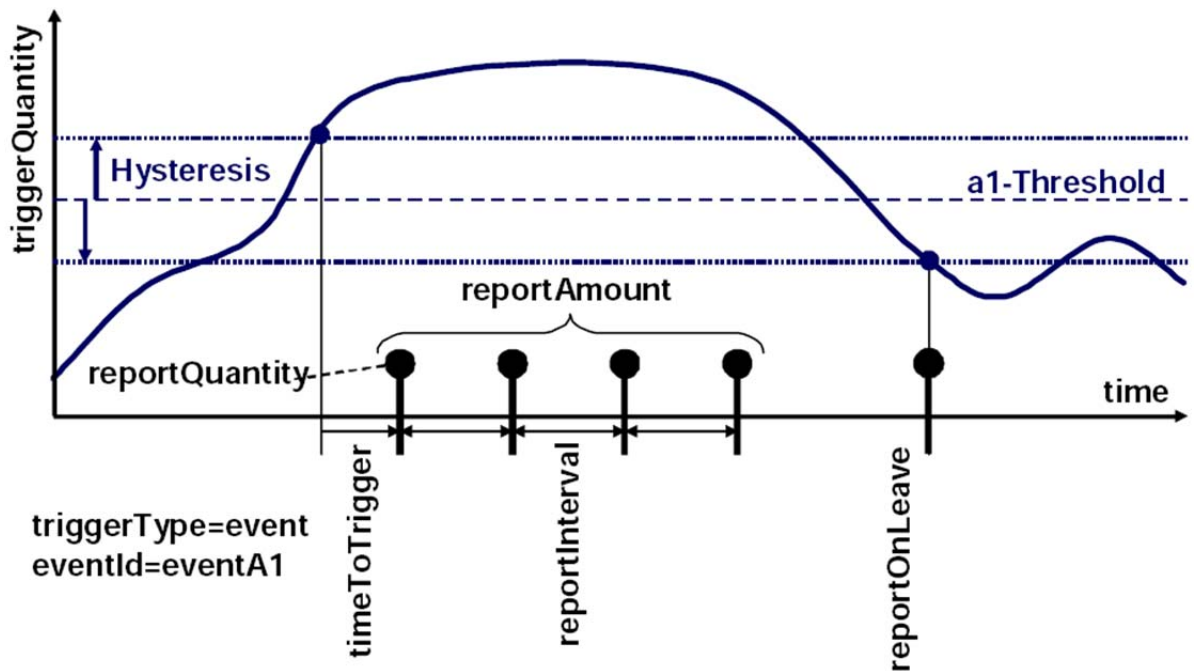
Figure 2.11: Event A1 based measurement reporting
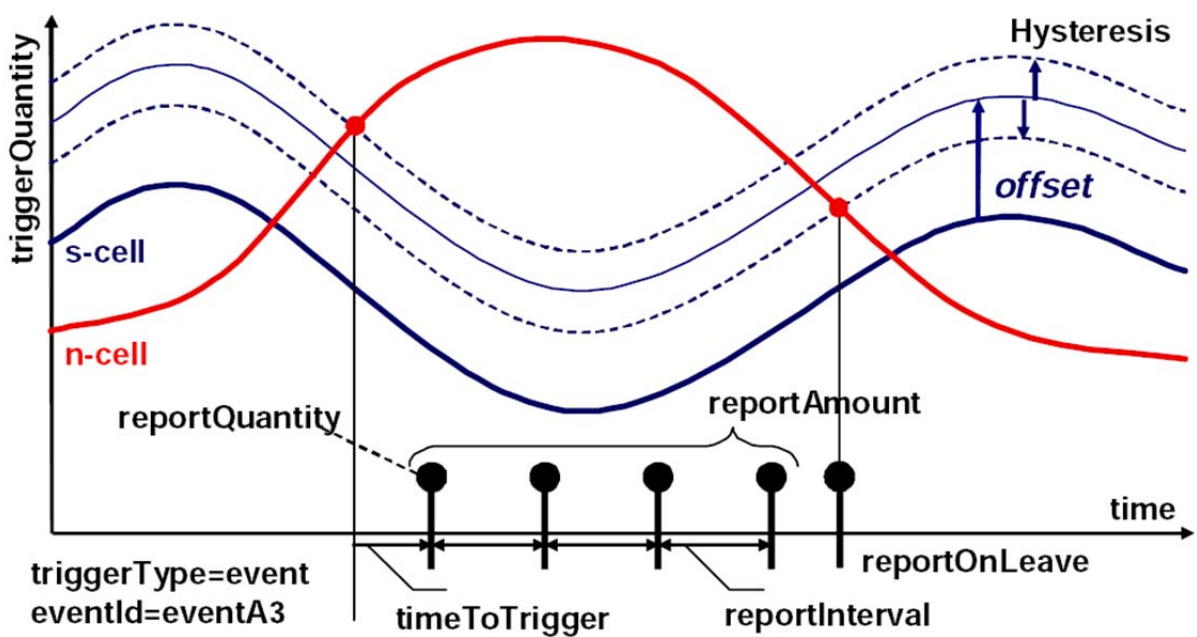
Figure 2.12: Event A3 based measurement reporting

34

2. *Periodical:* The UE keeps on sending the measurement reports one after another at a certain interval until it reaches a maximum number of measurement reports. The network specifies the interval between measurement reports and the maximum number of measurement reports to be sent.
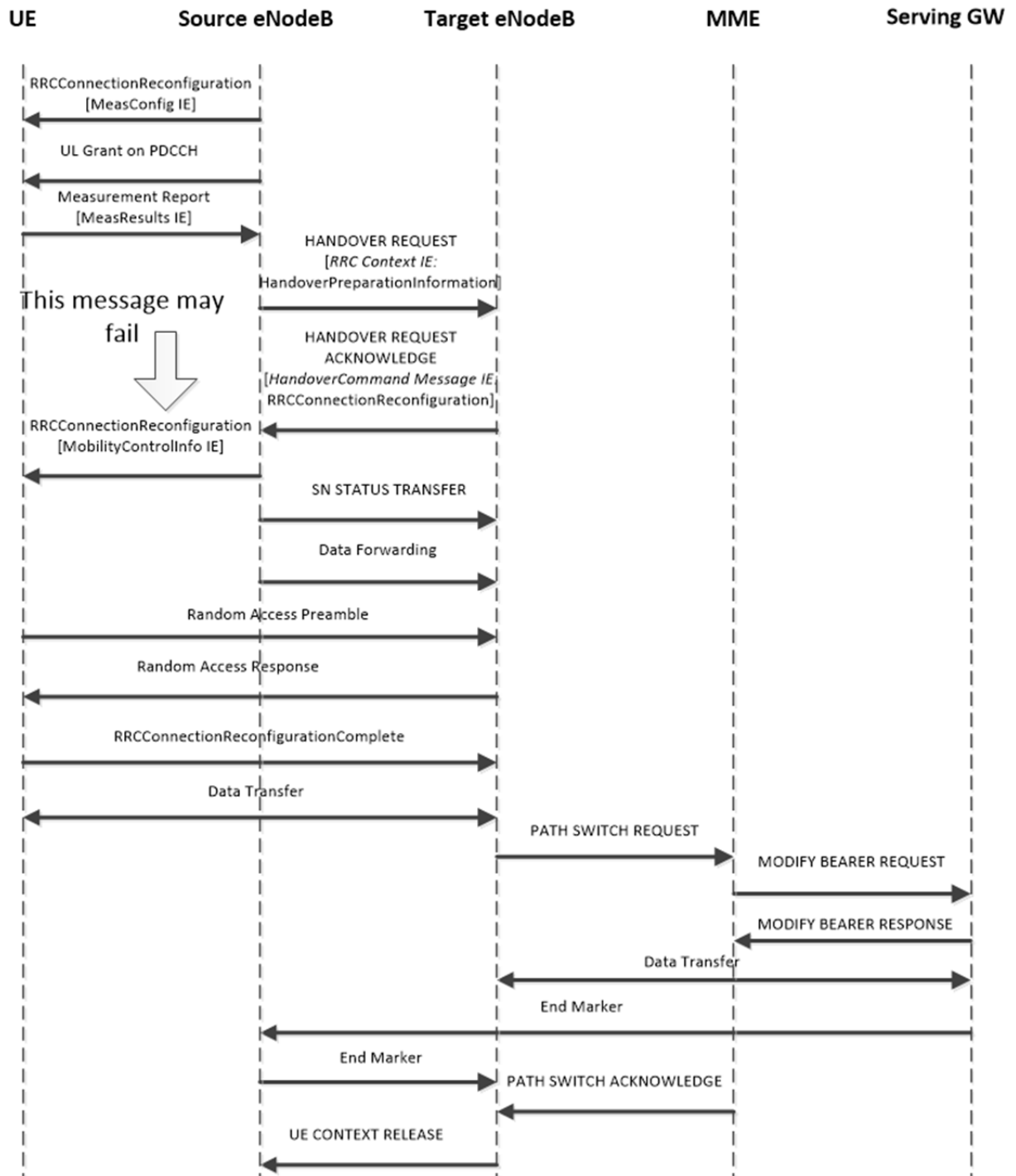
Figure 2.13: Exchange of messages during handover

When a handover is triggered, the following steps take place [14]. The whole handover procedure is shown in Fig 2.13 [1].

1. *Handover Preparation:* The source eNodeB sends a HANDOVER REQUEST message to the target eNodeB. If the target eNodeB is ready for allocation of the resources, then it sends the HANDOVER REQUEST ACKNOWLEDGE message to the source eNodeB.

2. *Handover Execution:* The source eNodeB sends the RRCConnectionReconfiguration message to the UE as an instruction for handover action. Earlier, this target fully created this RRCConnectionReconfiguration message and forwarded to the source eNodeB. This message includes the MobilityControlInfo IE which specifies how the UE would act at the target eNodeB. In order to avoid loss of data, the source eNodeB forwards data to the target eNodeB over X2 interface temporarily during handover action until data transfer begins between the UE and the target eNodeB. Then the UE performs downlink synchronization with the target eNodeB. The UE sends RRCConnectionReconfigurationComplete message to the target eNodeB notifying that it has accessed the target cell successfully.

3. *Handover Completion***:** The data transfer begins between the UE and the target eNodeB. The Serving Gateway switches the downlink data path from the source eNodeB to the target eNodeB. The target eNodeB sends UE CONTEXT RELEASE message to the source eNodeB and then the source eNodeB releases its resources. This completes the whole handover procedure.

The handover is required to take place on time. A too early or too late handover is undesired as shown below.

1. *Handover Too Late:* As the UE moves, if the handover occurs too late, the radio link with the source cell can get too poor. This can cause radio link failure (RLF) at the source cell. The RLF can lead to a handover failure (HF). In this case, typically because of RLF, the UE fails to receive RRCConnectionReconfiguration message from the source eNodeB, which is shown above in the handover execution phase. This failure of transmission of RRCConnectionReconfiguration message is indicated in Fig. 2.13. Thus, the UE cannot switch to the target eNodeB and HF occurs.

The HF can cause a call or session drop if it cannot be recovered. After a HF occurs, the UE can attempt to reestablish RRC connection with the strongest cell. A call or session drop occurs if the RRC connection reestablishment procedure fails. If the RRC connection reestablishment with the target cell is successful, the target cell can send a RLF Indication message to the source cell.

2. *Handover Too Early:* As the UE moves, if the handover occurs too late, the radio link with the target cell may not have become sufficiently strong yet. In this case, during or after the handover procedure, RLF can occur at the target cell and the UE tries to reestablish RRC connection with the source cell. After the RRC connection reestablishment, the source cell can send a RLF Indication message to the target cell. To inform the source cell about the failure, the target cell can send a Handover Report message setting 'Handover Report Type' as 'Handover Too Early'.

2.8.3 Scaling of Parameters for Speedy Users

The users moving at high speed may be subject to very poor cell edge SINR due to what is often called a dragging effect. This is because the cell switching is not triggered until the required conditions have been fulfilled for a certain time. This switching delay would let the UE move too far away from the old serving cell before switching to a new cell. The handover is then triggered in the dragging handover region shown in Fig. 2.14, in which the cell coverage is shown in dBm [15]. This results in very poor signal strength experience from the serving cell. It degrades the cell edge performance causing extremely low data rate. In the worst case, the dragging effect can cause RLF and HF, which are explained in Section 2.8.1 and 2.8.2.
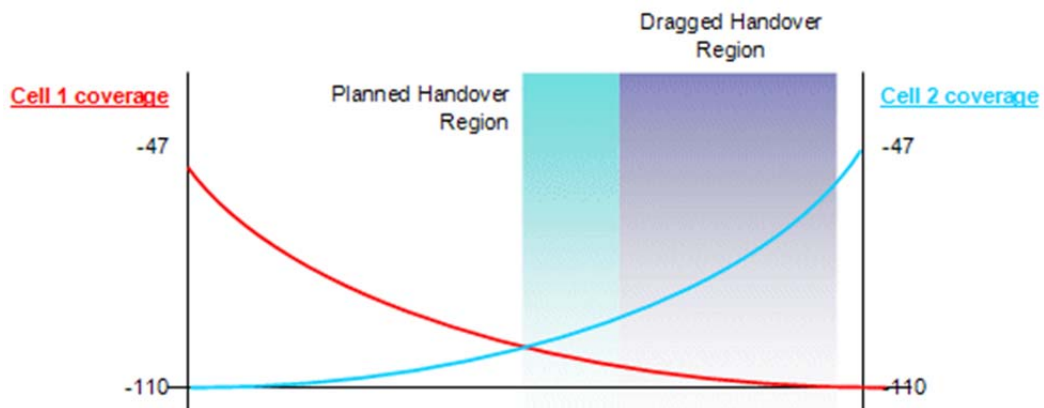
Figure 2.14: Shifting of the location of handover because of dragging effect

Both the eNodeB and the UE, typically, experience the lowest received signal power right before the cell switching. Let us denote this lowest received signal power as P_{R_Min}. The good quality of radio link can be maintained for a speedy UE and RLF and HF can be avoided, if the UE can switch to the new cell more quickly. This can be achieved by allowing less stringent cell switching conditions. The higher the speed of the UE is, the less stringent the cell switching conditions should be. In other words, P_{R_Min} should be made velocity independent as much as possible. For this purpose, 3GPP specifications allow scaling of cell switching parameters (e.g. reduction of hysteresis margin) depending on the user velocity. For this purpose, apart from the normal-mobility state supporting UEs at a low speed, the high-mobility state and medium-mobility state are defined as follows in existing 3GPP standards,. The RRC layer of the UE applies scaling of cell switching parameters in high and medium mobility state.

- *High Mobility State:* The UE enters high mobility state if there are more than N_{CR_H} number of cell reselections or handovers in last T_{CRmax} duration.
- *Medium Mobility State:* The UE enters medium mobility state if there are more than N_{CR_M} number but equal to or less than N_{CR_H} number of cell reselections or handovers in last T_{CRmax} duration.
- *Normal Mobility State:* The UE enters normal mobility state from medium or high mobility state if neither medium nor high mobility state is detected during $T_{CRmaxHyst}$. No scaling is performed in this state.

The RRC layer of the eNodeB configures the values of T_{CRmax}, N_{CR_H}, N_{CR_M} and $T_{CRmaxHyst}$ for both cell reselection and handover. For cell reselection, the RRC layer of the UE scales Q_{hyst} and $T_{reselectionRAT}$ as shown below.

- Q_{hyst}: In high or medium mobility state, Q_{hyst} is decreased by adding SF-High or SF-Medium, respectively, to Q_{hyst}. The eNodeB sends these SF-High and SF-Medium values on Q-HystSF IE via system information message and their values can be -6, -4, -2 or 0 dB.
- $T_{reselectionRAT}$: In high or medium mobility state, $T_{reselectionRAT}$ is decreased by multiplying by SF-High or SF-Medium, respectively. The eNodeB sends these SF-High and SF-Medium values on SpeedStateScaleFactors (SSSF) IE via system information message and their values can be 0.25, 0.5, 0.75 or 1.

For handover, the hysteresis is not scaled in the existing specification. In high or medium mobility state, *TimeToTrigger* is decreased by multiplying by SF-High or SF-Medium, respectively. The eNodeB sends SF-High and SF-Medium values on SSSF IE via RRCConnectionReconfiguration message.

2.9   Conclusion

The underlying technologies of LTE are too broad. This chapter discusses some fundamental concepts of LTE very briefly. It also introduces a few popular LTE-Advanced features. The topics were carefully selected for description such that they relate to the actual research work presented in this thesis. Thus, the discussions in this chapter can greatly help clarify different aspects presented in the following chapters. Since LTE is a new technology, a discussion on the basic concepts of LTE should be incorporated and so, it is provided in the next chapter. The discussion focuses on AS related aspects.

Each of the different layers of AS is assigned various operations. The basic AS operations are explained in the next chapter because many of these operations are the focus of the research presented in the subsequent chapters.

Chapter 3

Operation of Access Strata (AS)

The access strata (AS) in the protocol stack functions between radio network and user equipment and it is responsible for transporting data over the wireless connection and managing radio resources. The AS is comprised of several layers in LTE as shown in Section 2.3. This chapter provides a brief description of the operation of different layers of AS. A detailed description is available at [1].

3.1   Introduction

The control plane, the RRC is the top layer of Access Strata (AS) and it oversees the overall functions of AS. The RRC layer controls its lower layers in the AS as shown in Fig. 3.1 [16]. The RRC configures various parameters for layer 2 and layer 1 based on feedback from them.
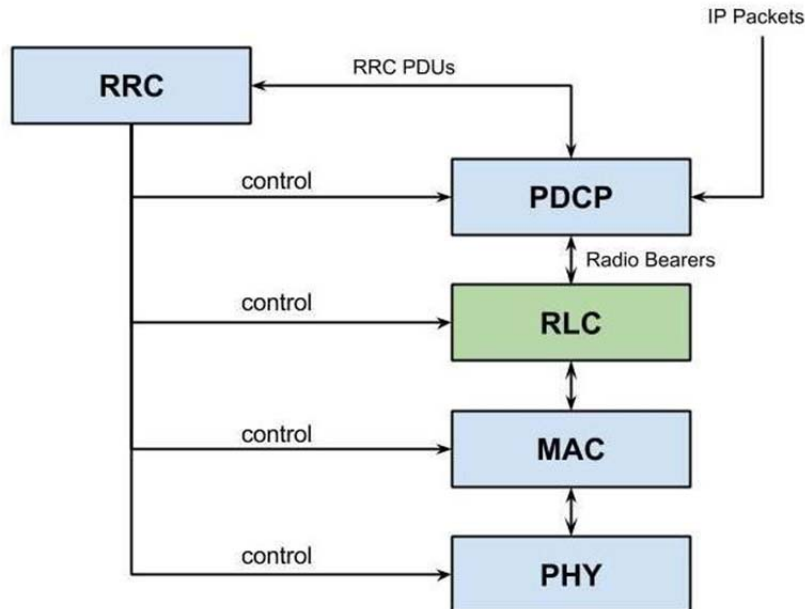
Figure 3.1: Control of RRC layer over its lower layers

In the user plane, PDCP first receives the IP packets as shown in Fig. 3.2 [17]. On the transmitting side, a layer receives the data packet as Service Data Unit (SDU), performs necessary processing and generates Protocol Data Unit (PDU). Then the layer sends the PDU

to its next lower layer and the lower layer receives it as its own SDU. Thus, the PDU of a layer becomes the SDU of the next lower layer. It occurs in the reverse way on the receiving side. A layer receives the data packet as a PDU, performs necessary processing, generates SDU and sends the SDU to its next upper layer.
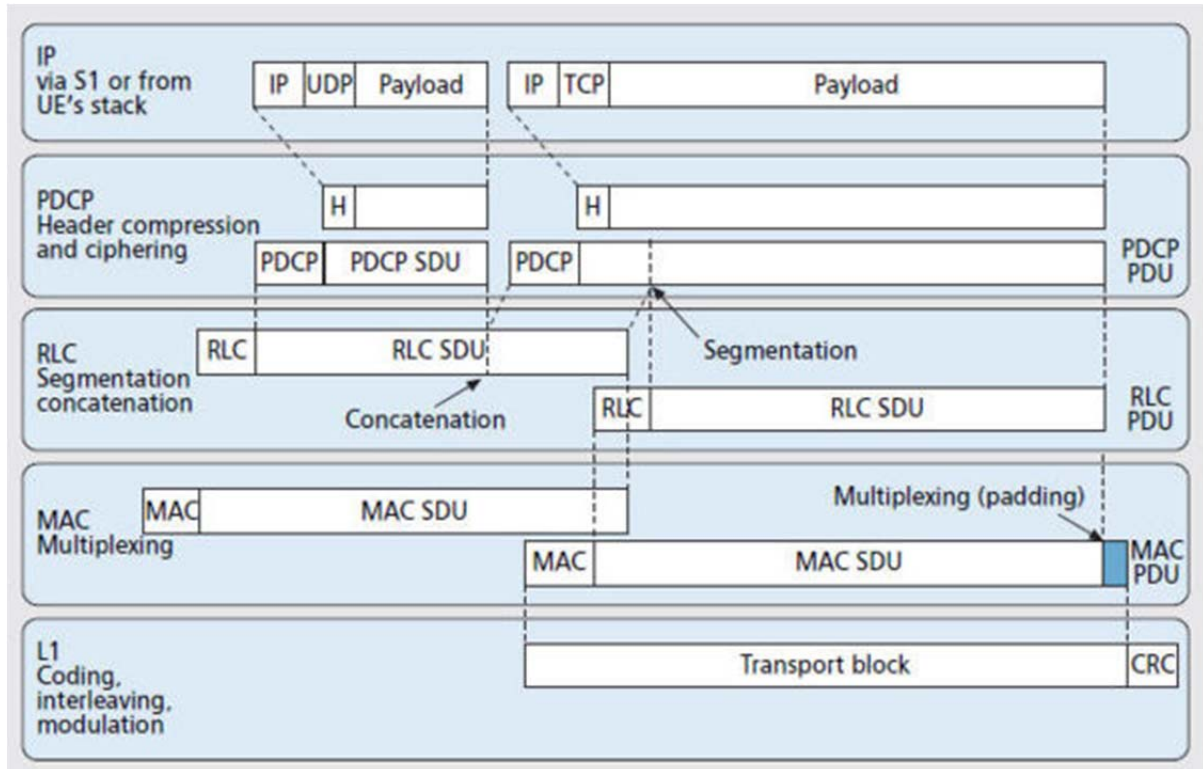
Figure 3.2: Processing of data packets in different layers in AS

3.2   Radio Resource Control (RRC)

The RRC layer oversees the operations of layer 1 and layer 2 in the Access Strata (AS). The operation of RRC layer is described in [1], [18]. The connection between the RRC layers of the UE and the eNodeB is referred to as RRC connection. When they are not connected, the UE is in RRC_IDLE state. Conversely, if a connection between them has been established, the UE is in RRC_CONNECTED state. A message generated by the RRC layer is called RRC message or layer 3 message. The layer 3 message contains various configuration parameters for layer 1 and layer 2. The RRC layer performs the following major functions.

− The RRC layer at the eNodeB broadcasts system information.
− The RRC layer at the eNodeB performs paging.

- The RRC layer at the eNodeB performs radio resource scheduling.

- The RRC layer at the eNodeB performs security activation in AS layer.

- The RRC layer at the eNodeB configures layer 1 and layer 2 parameters.

- The RRC layer at the eNodeB configures measurement procedure of the UE including measurement gaps and it conducts handover.

- The RRC layer at the UE performs cell selection.

- The RRC layer at the UE performs cell reselection.

- The RRC layer at the eNodeB and the UE performs establishment, modification and release of RRC connection.

- The RRC layer at the UE performs recovery from radio link failure.

3.3   Packet Data Convergence Protocol (PDCP)

The Packet Data Convergence Protocol (PDCP) is the top sublayer within layer 2 of AS. The operation of PDCP layer is described in [1], [19]. The PDCP layer performs the following major functions as shown in Fig. 3.3 [19].

1. *Processing User Plane and Control Plane Data Packet:* The PDCP layer contains PDCP entities. On the transmitting side, each PDCP entity receives PDCP SDUs from upper layers. It then performs addition of header and other necessary processing and generates PDCP PDUs. The PDCP entity passes the PDCP PDUs on to associated RLC entities. On the receiving side, each PDCP entity receives PDCP PDUs from RLC entities. Then it removes the header and performs other necessary processing and retrieves PDCP SDUs. Thereafter, it passes the PDCP SDUs on to upper layers. The PDCP entities are associated with RLC entities as follows.

    - RLC Unacknowledged Mode (UM): The RLC UM provides a unidirectional data transfer. Therefore, two RLC entities are used for a bidirectional bearer, a transmitting UM RLC entity and a receiving UM RLC entity. In this case, each PDCP entity is associated with two RLC entities, one for transmission and one for reception.

    - RLC Acknowledged Mode (AM): The RLC AM provides a bidirectional data transfer. A single AM RLC entity is used for a bidirectional bearer and it has two sides, a transmitting side and a receiving side. Thus, in this case, each PDCP entity is associated with one RLC entity.

2. *Security Function:* PDCP applies ciphering to the RRC PDUs and user plane data. It also applies integrity protection to the RRC PDUs.

3. *Header Compression and Decompression for User Plane Data:* The PDCP layer performs header compression of IP data at the transmitter side and header decompression of IP data at the receiver side in order to minimize protocol overheads transferred. The header compression also reduces transmission delay and packet loss rate. For this purpose, PDCP uses only RObust Header Compression (ROHC) protocol defined by the IETF (Internet Engineering Task Force) which is a robust and efficient technique.
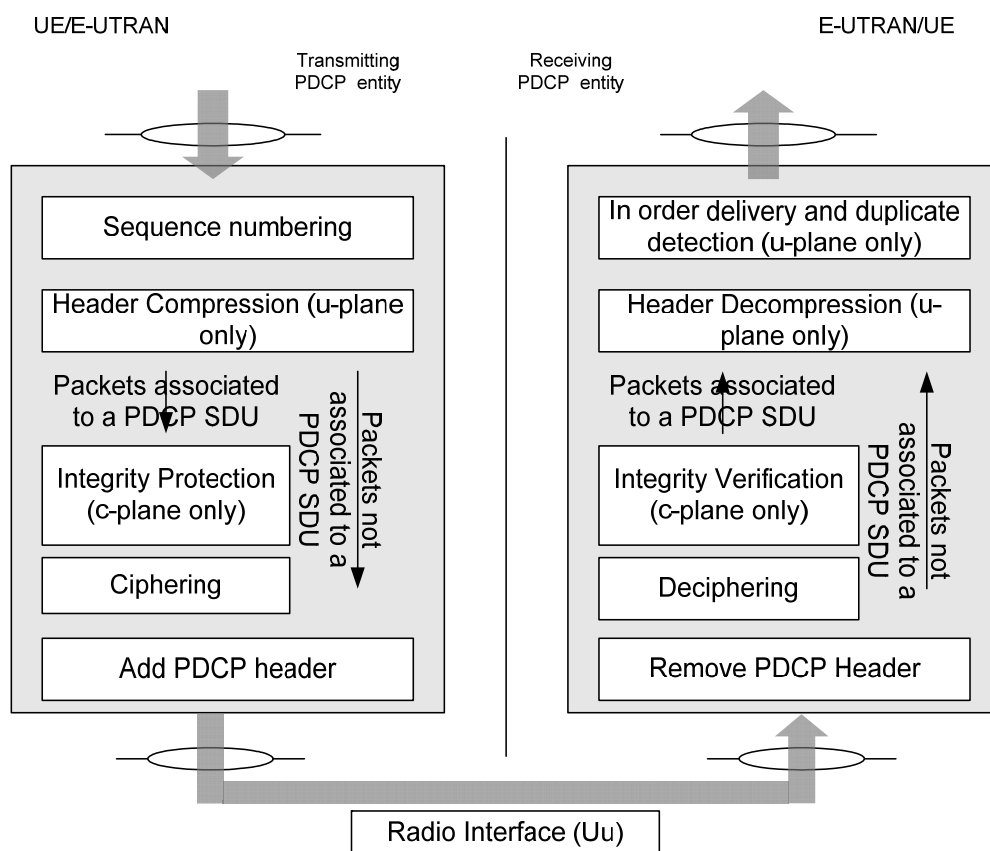
Figure 3.3: Operation of PDCP peer entities

4. *Discard of Data Packets in Queue:* In order to prevent from too much buffering at the transmitter when the data rate over the radio link gets worse, the PDCP layer can be configured to discard data packets after they have waited for transmission for certain time. This loss of packets is taken care of by upper layer protocols, TCP or RTP.

5. *Proper Delivery of Packets after Handover and Temporary Radio Link Failure:* The PDCP layer is reestablished after handover and when RRC connection re-

establishment occurs after temporary radio link failure. In this case, RLC and MAC layers are reset. Since the RLC layer is reset, it can no longer deliver PDCP SDUs in order. Also, there can be data missing due to interruption of service. It is then the PDCP layer that offers in-order delivery of packets, duplicate detection and necessary retransmission.

3.4   Radio Link Control (RLC)

The Radio Link Control (RLC) is the middle sublayer within layer 2 of AS. The operation of RLC layer is described in [1], [20]. The MAC layer notifies RLC when there is a transmission opportunity. It also notifies the possible size and number for RLC PDUs to be generated based on the availability of transport blocks. The transport block size depends on the resources available and Modulation and Coding Scheme (MCS) to be used on the basis of physical channel conditions. The transmitting and receiving RLC entities work as peer entities for data transfer in both uplink and downlink. The RLC entity is configured in any of the following three modes.

1.   Transparent Mode (TM)

2.   Unacknowledged Mode (UM)

3.   Acknowledged Mode (AM)

3.4.1   Transparent Mode (TM)

The Transparent Mode (TM) is used transfer system information messages, paging messages and certain RRC messages. The TM is not used for transfer of any user data. The TM provides a unidirectional data transfer and it does not guarantee the delivery the information. The TM transfers SDUs without any processing. Acknowledged Mode (AM)

3.4.2   Unacknowledged Mode (UM)

The Unacknowledged Mode (UM) does not guarantee the delivery the information and packet losses propagate to higher layers. It is suitable for delay sensitive and error tolerant real time (RT) applications. Thus, voice, live streaming video, online gaming, etc. use UM. The UM is not used for signaling messages. The UM performs in-sequence delivery, concatenation, segmentation, reassembly and discarding duplicate during transfer of SDUs.

The UM provides a unidirectional data transfer. Two RLC entities are used for a bearer, a transmitting UM RLC entity and a receiving UM RLC entity as shown in Fig. 3.4. The transmitting UM RLC entity performs the following functions.

    i.   The transmitting UM RLC entity receives RLC SDUs from PDCP layer and it performs segmentation and/or concatenation of RLC SDUs, if necessary.

    ii.   It adds RLC Header and produce RLC PDU.

    iii.   It passes RLC PDUs to MAC layer for transmission over the air.

The receiving UM RLC entity performs the following functions.

    i.   The receiving UM RLC entity receives RLC PDUs from MAC layer and puts them in reception buffer.

    ii.   It reorders out of sequence RLC PDUs.

    iii.   It detects lost RLC PDUs.

    iv.   It discards duplicate RLC PDUs. If there is any duplicate RLC PDUs, the receiving RLC entity detects them based on the sequence number and discards them. It may happen that the feedback is decoded incorrectly in the HARQ process. If an ACK is misinterpreted as NACK, then the correctly received transport block can be retransmitted and this can create a duplicate RLC PDU.

    v.   It removes RLC Header and reassembles complete RLC SDUs. The reassembly is the reverse process of the segmentation and concatenation. If any segment of a RLC SDU is missing, the whole RLC SDU is discarded. It also discards any RLC PDUs which have been received but cannot be reassembled into RLC SDUs.

    vi.   It passes complete RLC SDUs on to PDCP layer.

The receiving RLC entity may receive out of sequence RLC PDUs. Since, multiple HARQ processes run in the MAC layer, there will be extra delays for transport blocks which are retransmitted in the HARQ protocol. This is the key reason for the possibility of receiving out of sequence RLC PDUs. Besides, different propagation delays can make the PDUs out of sequence. The RLC PDU includes Sequence Number (SN) field in its header which indicates the sequence number of the RLC PDU. The transmitting UM RLC entity increases the sequence number by one in every successive PDU. The receiving RLC entity reorders the RLC PDUs based on their sequence numbers. If the receiving RLC entity finds a gap in the sequence of the received RLC PDUs, it assumes that the missing RLC PDU has been delayed due to HARQ retransmission of the corresponding transport block. So, it then starts a

reordering timer. It stores out of sequence RLC PDUs, does not deliver them to the higher layer and waits for the missing RLC PDU until the timer expires. If the timer expires without arrival of the missing RLC PDU, the receiving RLC entity infers that HARQ has failed to recover the RLC PDU and the missing RLC PDU has been lost. It then reassembles only the RLC PDUs which can generate complete RLC SDUs. This precludes the receiving RLC entity from waiting for missing RLC PDUs indefinitely.
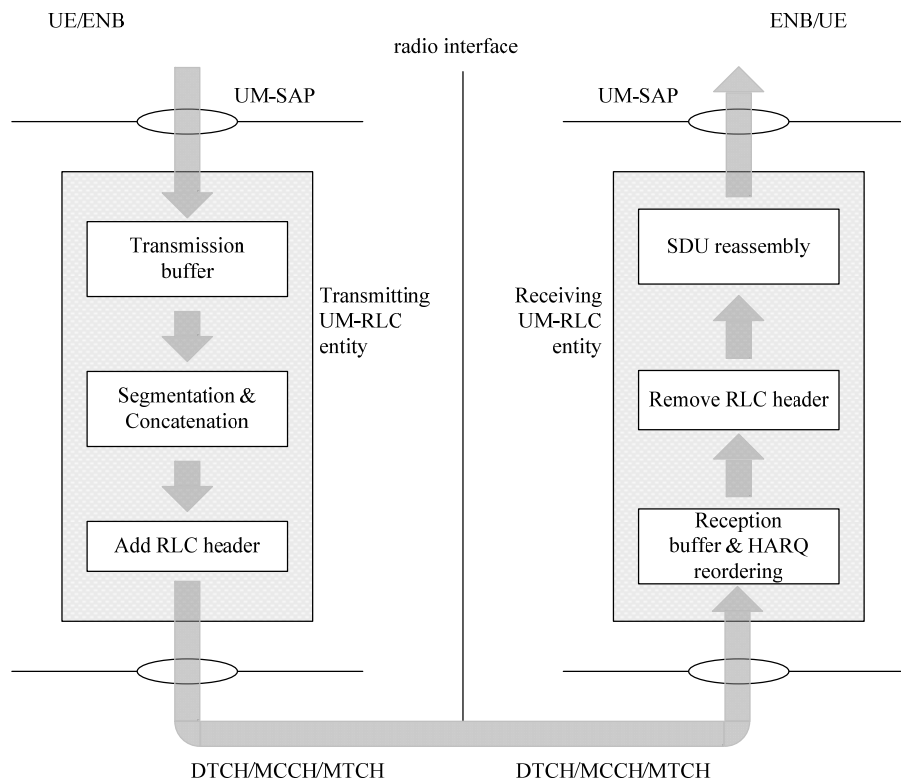
Figure 3.4: Operation of peer entities in Unacknowledged Mode (UM)

3.4.3   Acknowledged Mode (AM)

The RRC layer at eNodeB notifies using RLC-Config IE whether Unacknowledged Mode (UM) or Acknowledged Mode (AM) would be used for data transfer. The Acknowledged Mode (AM) provides reliable in sequence delivery of data using Automatic Repeat Request (ARQ). The ARQ is an error control method for data transmission to achieve reliable data transmission. The receiver sends an indication if it has correctly received the PDU or not. If the packet is not received, the transmitter retransmits the PDU. Thus, in addition to in-sequence delivery, concatenation, segmentation, reassembly and discarding duplicate, the AM also performs error correction using Automatic Repeat Request (ARQ) and

resegmentation of PDUs, if required in retransmission. The same sequence number and reception buffer are used for both ARQ operation in RLC layer and HARQ operation in MAC layer. Thus, the HARQ reordering does not need additional sequence number and reception buffer.

Because of the use of ARQ, acknowledged mode is suitable for error sensitive and delay tolerant applications. So, it is most commonly used for data transfer, especially for interactive and background type applications, e.g. HTTP and FTP transfer. It is particularly suitable for TCP based traffic. Thus, buffered streaming video, web browsing, e-mail, chat, FTP, and P2P file sharing, use AM. Streaming applications may also use AM RLC if they are not too delay intolerant. The AM RLC is also used for transfer most of the RRC messages.
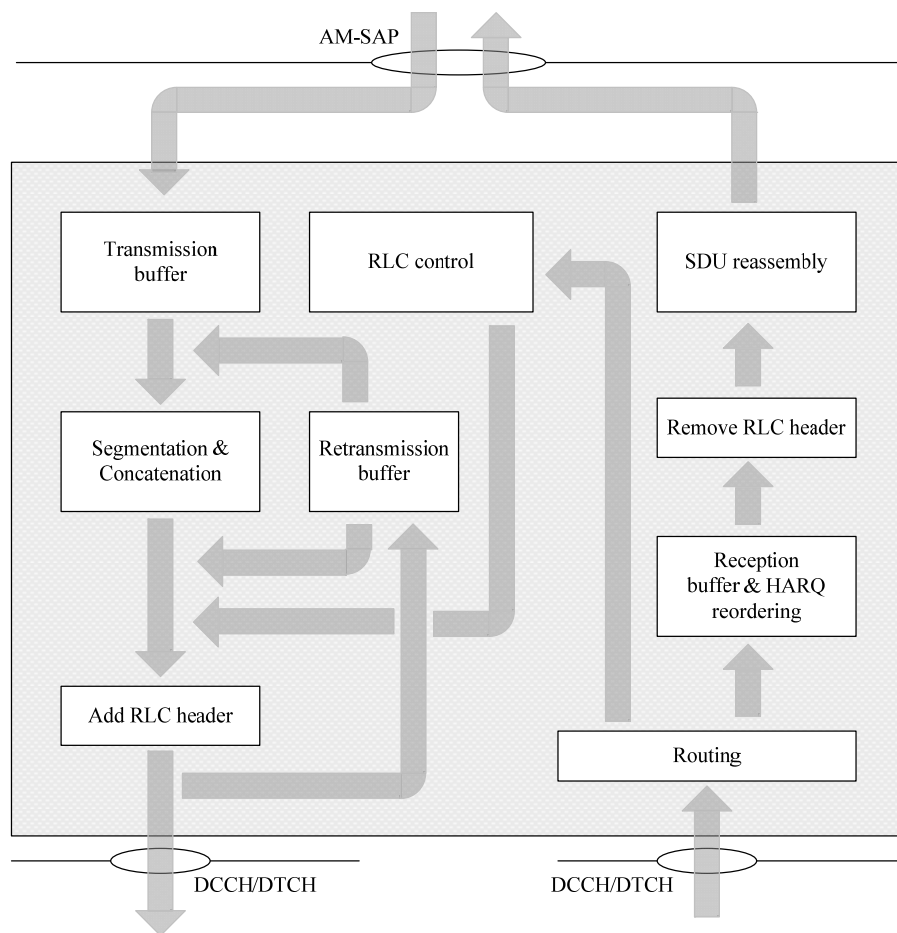
Figure 3.5: Operation of peer entities in Acknowledged Mode (AM)

The AM provides a bidirectional data transfer. A single AM RLC entity is used for a bearer and it has two sides, a transmitting side and a receiving side as shown in Fig. 3.5. Both transmitting side and receiving side use the sliding window for ARQ protocol. The sliding

window in the transmitting side covers the RLC PDUs which are not sent as well as the RLC PDUs which are sent but not yet acknowledged. The sliding window in the receiving side covers expected RLC PDUs. All RLC PDUs may not be received with sequentially higher numbers and they may rather be intermittent, i.e. some RLC PDUs may be missing while RLC PDUs with higher sequence numbers are received. Then the sliding window is moved ahead of the sequentially received RLC PDUs and it covers intermittent RLC PDUs.

In case of transmission of the original packets, the transmitting side of the AM RLC entity performs the following functions.

i. The transmitting side of the AM RLC entity receives RLC SDUs from upper layers and puts RLC SDUs in transmission buffer.

ii. It performs segmentation and/or concatenation of RLC SDUs, if necessary.

iii. It makes a copy of the transmit buffer for possible retransmissions.

iv. It adds RLC Header and produce RLC PDU.

v. It passes RLC PDUs to MAC layer for transmission over the air.

In case of reception of the original packets, the receiving side of the AM RLC entity performs the following functions.

i. The receiving side of the AM RLC entity receives RLC PDUs from MAC layer.

ii. It reorders out of sequence RLC PDUs. Like UM mode, if the reordering timer expires without arrival of the missing RLC PDU, the receiving RLC entity infers that HARQ has failed to recover the RLC PDU and the missing RLC PDU has been lost. However, at this juncture, unlike UM mode, it does not reassemble the RLC PDUs to generate RLC SDUs. It rather sends a status report.

iii. It detects lost RLC PDUs.

iv. It discards duplicate RLC PDUs.

v. If a RLC PDU is received correctly, it is marked accordingly.

vi. It moves the sliding window ahead of the sequentially received RLC PDUs with no missing PDUs in between. The sliding window covers intermittently received RLC PDUs.

vii. For sequentially received RLC PDUs, it removes RLC header and reassembles complete RLC SDUs. The reassembly is the reverse process of the segmentation and concatenation.

viii. It passes complete RLC SDUs on to upper layers.

ix. It sends status report to the transmitting side of the AM RLC entity when needed. The status report includes positive and negative acknowledgements of reception.

If the transmitting side of the AM RLC entity receives positive acknowledgement on status report, then it performs the following functions.

i. It moves its sliding window ahead of the acknowledged RLC PDUs. The sliding window covers the RLC PDUs which are not acknowledged whether they are sent or not.

ii. It removes the acknowledged data from the retransmission buffer.

If the transmitting side of the AM RLC entity receives negative acknowledgement on status report, then it performs the following functions.

i. It retrieves the negatively acknowledged data from the retransmission buffer.

ii. It retransmits the negatively acknowledged data. If the available resource for retransmission is smaller than the original RLC PDU size, then the RLC PDU is re-segmented for retransmission.

The receiving side of the AM RLC entity performs the following functions after reception of the retransmitted data from the peer transmitting side of the AM RLC entity. .

i. The receiving side of the AM RLC entity updates its buffer with retransmitted RLC PDUs. It moves the sliding window ahead of the sequentially received RLC PDUs with no missing PDUs in between. The sliding window covers intermittently received RLC PDUs, if there is still any.

ii. For sequentially received RLC PDUs, it removes RLC header and reassembles complete RLC SDUs.

iii. It passes complete RLC SDUs on to RRC layer or PDCP layer.

The retransmission of a RLC PDU or a portion of a RLC PDU recurs if needed. However, the transmitting side of the AM RLC stops retransmission when it reaches a maximum limit. This maximum limit is specified by the RRC layer at eNodeB. When the maximum number of retransmissions is reached, the RLC layer reports it to RRC layer and the RRC layer determines that a radio link failure has occurred.

The reordering timer value is configured the same way in the case of AM and UM. The possible reordering timer values are also same. The RRC layer at eNodeB configures the

reordering timer value. The eNodeB directly applies the reordering timer value for uplink data transfer. The eNodeB notifies the UE of the reordering timer value for downlink data transfer using the RLC-Config IE. The reordering timer value can be any duration between 0 ms and 100 ms with 5 ms gaps or any duration between 100 ms and 200 ms with 10 ms gaps [18].

3.5   Medium Access Control

The Medium Access Control (MAC) is the bottom sublayer within layer 2 of AS. The operation of MAC layer is described in [1], [21]. The MAC layer performs the following major functions.

1. *Multiplexing, Demultiplexing and Mapping of Data Packets:* The MAC layer treats RLC Protocol Data Units (PDUs) from different logical channels as MAC Service Data Unit (PDUs), performs necessary processing, generates MAC PDUs and delivers MAC PDUs onto downlink or uplink transport channels as Transport Blocks (TBs). A MAC PDU is fitted into a transport block and the transport block does not contain more than one MAC PDU. There may be one or more MAC control elements in the MAC PDU. Also, padding bits are added at the end of the MAC PDU, if it is needed to complete the transport block size. A MAC PDU is transmitted in a subframe, which is 10 ms long. However, more than one transport block can be transmitted in a subframe when multiple antenna techniques are used. The MAC layer protects each transport block with a CRC.

2. *Discontinuous Reception (DRX):* The MAC layer of the UE performs DRX, which is explained in Section 3.5.1.

3. *Scheduling of Radio Resources among UEs:* The scheduler in the MAC layer of the eNodeB allocates the available radio resources among different UEs for both uplink and downlink transmission in a cell. This is explained in Section 3.5.2.

4. *Random Access Procedure:* The UE attempts to gain access to the network and create a scope for communication with the network using random access procedure. The MAC layer performs the random access procedure and it establishes layer 1 connectivity between the UE and the eNodeB. The random access establishes uplink synchronization for the UE and also, it allocates resources for uplink transmission.

5. *Uplink Time Alignment:* The physical layer of the UE is required to have uplink time alignment or synchronization with the eNodeB so that there is no interference among

the uplink transmissions from different UEs. This requires consideration of downlink and uplink propagation delays. A timing control is required to establish uplink time alignment as well as to update it constantly. The MAC layer at the eNodeB performs this timing control.

6. *Hybrid ARQ (HARQ) Operation:* Forward error correction (FEC) attempts to deliver correct data packets but when it fails, the incorrectly received packets need to be retransmitted. In case of automatic repeat request (ARQ), the receiver sends an acknowledgment to indicate whether it has correctly received the packet or not. HARQ combines FEC and ARQ, to minimize the number of retransmissions. In this case, when the receiver fails to decode the data block, it sends a NACK to the transmitter, but it keeps bits from the failed attempt for future use. If a NACK is received or a certain time elapses without any feedback, the transmitter retransmits the data block. Then the receiver HARQ soft-combines bits from the previous attempt with the current retransmission.

7. *Transport Format Selection:* The MAC layer of the eNodeB is responsible for transport format selection in both uplink and downlink. The transport format includes the Modulation and Coding Scheme (MCS) and Transport Block Size (TBS) to be used for the transport block. The eNodeB indicates the transport format for both uplink and down transport blocks using PDCCH.

8. *Buffer Status Reporting:* If there is already PUSCH resources allocated for uplink transmission, the UE sends Buffer Status Reports (BSRs) on PUSCH in order to report pending data in its uplink buffers. For this purpose, the MAC layer of the UE includes BSR MAC control element in the MAC PDU.

9. *Sending Scheduling Request***:** When there is no ongoing uplink data transfer on PUSCH, the UE sends Scheduling Request (SR) on PUCCH seeking resources provided it has PUCCH resource allocation. An SR may also be sent after the transmission of a regular BSR when PUSCH is allocated.

10. *Power Headroom Reporting:* The uplink power control considers the power transmit capability of the UE. It sets the uplink power level within its maximum uplink transmission power. The UE sends power headroom report (PHR) to notify the eNodeB how much more or less power it is capable of transmitting. For this purpose, the MAC layer of the UE includes Power Headroom MAC control element in the MAC PDU. The power headroom reporting also assists the eNodeB to allocate uplink resources among different UEs appropriately.

3.5.1   *Discontinuous Reception (DRX)*

The incredibly higher data rates and better quality of services have motivated the advent and use of many different applications in LTE. This causes the battery of the user equipment (UE) to drain out quickly and the battery power saving has become a significant concern. The key technique to save power and prolong battery life is to let the UE switch off the receiver circuitry periodically. This is referred to as discontinuous reception (DRX) and it is carried out at MAC layer. DRX mechanism is explained in [1], [19]. The periodic cycle in DRX is called DRX cycle. The DRX cycle consists of an active period in the beginning, called on duration and then a period with the receiver circuitry off, called off duration. DRX can be applied in both RRC_IDLE state and RRC_CONNECTED state. In RRC_IDLE state, the UE wakes up once in every DRX cycle and looks for paging message. The active period of the DRX cycle has one Paging Occasion (PO), i.e. the UE has one PO per DRX cycle. Thus, the DRX cycle is the same as the paging cycle for a UE. The length of the DRX cycle in RRC_CONNECTED state is made, at the maximum, equal to the length of the DRX cycle in RRC_IDLE state. The maximum possible length of the DRX cycle is 256 radio frames for both RRC_IDLE state and RRC_CONNECTED state.

In RRC_CONNECTED state, the eNodeB can optionally configure DRX operation. The RRC layer at eNodeB configures DRX by sending DRX-Config IE. If DRX is not configured, the UE monitors PDCCH continuously in order to find out possible allocations for downlink or uplink transmission. When DRX is configured, the eNodeB specifies a DRX cycle. The UE monitors PDCCH during on duration and the UE pauses downlink reception during DRX period as shown in Fig. 3.6 [22]. The DRX-Config IE specifies the length of the on duration. When a packet arrives, it cannot be scheduled until the off period of the current DRX cycle is over. This enhances the packet delay, which is the main drawback of DRX operation. There are two types of DRX cycles, namely, short and long DRX cycles and as their names suggest, their lengths are short and long, respectively. The long DRX cycle allows increased power saving but it reduces the frequency of opportunities for scheduling via PDCCH and thus, it causes a longer packet delay. The short DRX cycles are optional and they are helpful when there are good chances of frequent data packets. They are particularly suitable when there are transmissions of small data at short but regular intervals, for example, VoIP. If the short cycle is configured, the UE initiates DRX with short cycles but it transitions to long cycles after a maximum number of short cycles. If the short DRX cycle is not configured, the UE initiates DRX operation with long DRX cycles. The DRX-Config IE

specifies the maximum number of short cycles, the length of short and long cycles and the length of on duration. Besides, the eNodeB specifies the particular radio frame and subframe in which the long or short cycle begins.

If the UE detects any packet scheduling during an on duration of DRX cycle, it stops DRX and begins continuous data transfer. When the data transfer stops, the UE starts an inactivity timer. The DRX-Config IE specifies the length of the inactivity timer. If new packets arrive, they can be scheduled while the inactivity timer is running and so, the UE keeps monitoring. If no scheduling is detected until the expiry of the timer, the UE initiates DRX cycles immediately. On the other hand, if any scheduling occurs while the inactivity timer is running, data transfer resumes and the inactivity timer restarts after this data transfer finishes. Thus, the inactivity timer can restart again and again. It is uncertain how many times the inactivity timer may restart as such and how long it would run before getting reset.
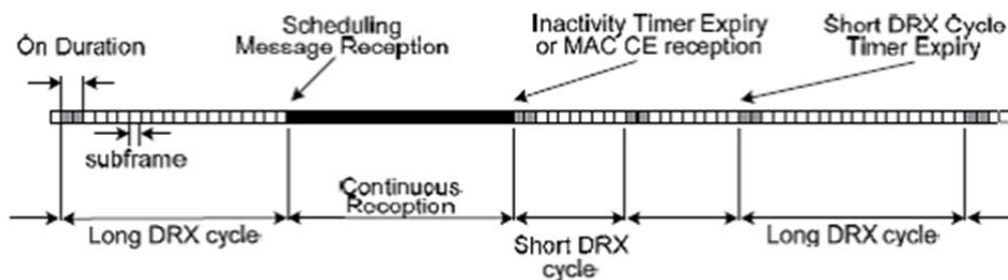
Figure 3.6: DRX operation in RRC_CONNECTED state

Alternatively, the eNodeB can send an explicit command by including DRX Command MAC Control Element in the MAC PDU during continuous reception. Then the UE initiates a short DRX cycle provided the short DRX cycle is configured. If the short DRX cycle is not configured, the UE initiates a long DRX cycle instead.

HARQ is independent of DRX operation. The UE wakes up from sleeping state to monitor PDCCH for possible HARQ retransmissions and ACK/NAK signaling on PHICH and also, to transmit HARQ feedback when they are due. Each HARQ process expects HARQ Round Trip Time (HARQ RTT) between transmission and retransmission of a transport block when the decoding fails. The HARQ RTT is estimated as 8 ms. Therefore, in case of downlink, the HARQ process at the UE maintains a HARQ RTT Timer to allow the UE to sleep during the HARQ RTT. The HARQ RTT Timer is set to 8 ms. If DRX operation is configured, then the UE pauses monitoring PDCCH for certain period. If decoding of a downlink transport block fails, the UE still pauses monitoring PDCCH but for HARQ RTT period assuming that the retransmission of the transport block might take place after this

HARQ RTT. Of course, the UE can pause monitoring PDCCH as such only if any other schedule does not come up for monitoring PDCCH. When the HARQ RTT Timer expires, the UE starts monitoring PDCCH for the retransmission. At this moment, the UE monitors PDCCH for a maximum period specified by DRX retransmission timer. The DRX-Config IE specifies the length of the DRX retransmission timer. The UE does not enter DRX when uplink HARQ retransmission is due.

3.5.2   *Scheduling of Radio Resources*

The scheduler in the MAC layer of the eNodeB allocates the available radio resources among different UEs and among the radio bearers of each UE for both uplink and downlink transmission in a cell through proper handling of priority. It is also possible that the MAC layer allocates different parts of the time-frequency resource grid in different parts of a cell.

In the time domain, the scheduling decision can be modified at every Transmission Time Interval (TTI) for a UE which is 1 ms or one subframe or two slots long. Thus, a pair of resource blocks which occupy the same set of 12 subcarriers and are contiguous in time is used as the granularity of scheduling decision. The number of transport blocks transmitted to the UE in every TTI is one if there is no spatial multiplexing and it can be two if spatial multiplexing is applied.

The PDCCH assigns resource blocks on PDSCH on the same subframe for downlink transmission. On the other hand, the uplink resource blocks on PUSCH are allocated 4 subframes after the transmission of the resource grant PDCCH. This time interval allows the UE to dequeue its data and prepare its transport block according to the specified attributes.

The MAC layer of eNodeB attempts to make appropriate apportionment of the resources with certain objectives as follows.

- Required QoS for applications.
- Optimized spectral efficiency ensuring high cell throughput under existing channel conditions.
- Fairness among UEs and applications.
- Limiting the impact of interference through special handling of cell edge users.
- Load balancing among cells.

The actual scheduling algorithm depends on the implementation at MAC layer of eNodeB. The scheduler may consider at least the following factors while making decisions in an attempt to make appropriate apportionment of the resources.

1. *Channel Dependent Scheduling (CDS):* The Channel Dependent Scheduling (CDS) can be made in both time and frequency domains. The scheduling adapts to channel variations and thus, it attains link adaptation. A user with better channel quality is given more resources as the user can make good use of these resources leading to higher cell throughput. Thus, the transmission occurs at fading peaks. The CDS requires the availability of sufficient information on uplink and downlink channel conditions. This information needs to be available for each parts of the bandwidth to allow CDS in frequency. The eNodeB can configure the availability of information with more or less granularity in frequency where a higher granularity consumes more resources for this information. Moreover, the eNodeB can configure frequency of the availability of the information in time. A more frequent availability represents the variation of radio channel better but at the cost of more resources for this information.

2. *Apportionment among UEs:* The channel dependent scheduling leads to higher cell throughput and on the other hand, the scheduling should maintain some fairness among the users in their resource allocations. Thus, there is a tradeoff between fairness and cell throughput. The scheduler can exercise various methods, for example, Round Robin (RR), Maximum C/I, Proportional Fair (PF), etc.

3. **Queue and Priority Aware Scheduling:** The scheduler performs queue-aware scheduling and it adapts to instantaneous traffic situation for dynamic sharing of the common resources among UEs as well as among logical channels.

4. **Modulation and Coding Scheme (MCS) and Transport Block Size (TBS):** The scheduling decision depends on the Modulation and Coding Scheme (MCS) used on the allocated resources to achieve certain data rate.

5. *Localized or Distributed Scheduling:* The scheduler employs either localized or distributed scheduling in downlink as shown below. The localized scheduling uses contiguous blocks of subcarriers. It attempts to make use of the best transmission band. Thus, it achieves Frequency Selective Scheduling (FSS). However, the localized scheduling may not provide good frequency diversity. Also, in this case, a low rate user may block required channel allocation for a high rate user, especially, if channel dependent scheduling is used. On the other hand, the distributed scheduling spreads the physical resource blocks across the subcarriers in frequency domain.

55

Thus, it achieves Frequency Diverse Scheduling (FDS). However, in this case, the channel estimation is degraded, especially, for very large repetition factors.

6. *Power or Bandwidth Limitation:* The scheduling decision accounts for power or bandwidth limitation for the particular scenario. The cell edge users may be power limited whereas the users close to the eNodeB may be bandwidth limited.

7. *Inter-Cell Interference Coordination (ICIC):* The scheduling depends on the implementation of ICIC among neighboring cells.

8. *Load Balancing among Neighboring Cells:* The scheduling depends on load balancing among neighboring cells.

3.6   Physical Layer

The physical layer is the bottom layer of AS. The operation of physical layer is described in [1], [23]−[25]. There can be one or two transport blocks in a TTI as shown in Section 3.5.2. A data transport block in the TTI, mapped on PDSCH or PUSCH, goes through the following general steps in physical layer processing.

1. *CRC Insertion:* A 24 bits long cyclic redundancy check (CRC) is computed over the transport block and it is added at the end of the transport block for error detection.

2. *Coding:* A rate 1/3 turbo encoder is used with quadratic permutation polynomial (QPP) interleaver. The code blocks size is 6144 bits at the maximum as it is the maximum size allowed by the turbo coder. If the number of bits becomes more than 6144 after addition of CRC, then the transport block is segmented into code blocks (CBs). The code blocks have fixed size of 6144 bits and for this size requirement, padding zero bits may be added at the beginning of the first code block. The code blocks are coded separately. This segmentation allows parallel coding and decoding and speeds up the process. Each code block includes its own CRC.

3. *Rate Matching:* The rate matching allows matching the number of bits to be transmitted exactly with the number of available resource elements. For this purpose, the rate matching applies puncturing and repetition on the rate 1/3 turbo encoder output. The rate matching attempts to maximize the number of new bits in HARQ retransmissions.

4. *Code Block Concatenation:* The code blocks are concatenated to obtain coded transport block.

5. *Scrambling:* The transport block is scrambled with length-31 Gold sequence. There are 2^{31} sequences as such and they are not cyclic shifts of each other. The scrambling sequence generator is re-initialized for each transport block and the initialization value is a function of the Physical Cell Identifier (PCI) of the cell and the Radio Network Temporary Identifier (RNTI) of the UE. This randomizes interference among cells and among UEs.

6. *Modulation:* The information bits are modulated with QPSK, 16-QAM or 64-QAM. PDCCH and PBCH use QPSK.

7. *Mapping to Transmission Layers:* Multiple streams of data are generated for mapping to different transmit antenna ports identified by transmission layers.

8. *Precoding:* Each transmission layer is precoded with certain matrix if MIMO is used.

9. *Mapping to Resource Elements:* The symbols are mapped to resource elements in the resource blocks. It uses frequency-first mapping, i.e. the symbols are mapped by incrementing the subcarrier index first for all the allocated resource blocks in a symbol period and thereafter, the next symbol period is taken.

10. *OFDM Modulation:* The symbols are fed to IFFT for OFDM modulation.

The data transfer in downlink involves a few basic steps as follows and it uses the physical channels shown in Section 2.5.4. Fig. 3.7 explains this transfer.

1. *Feedback from the UE:* The UE sends channel quality indicator (CQI), precoding matrix indicator (PMI), and rank indication (RI) as feedback to the eNodeB on PUCCH if there is no uplink data transfer and on PUSCH if there is ongoing uplink data transfer.

2. *Allocation of Resources:* Based on the feedback from the UE, the eNodeB allocates radio resources on PDSCH for downlink transmission. The eNodeB uses PDCCH signaling to inform the UE about the allocated resource blocks and configuration used in data transmission. The UE decodes PDCCH and obtains necessary configuration for the allocated resources. The configuration information includes Modulation and Coding Scheme (MCS) used, Transmitted Rank Indication (TRI) indicating the rank used for MIMO and Transmitted Precoding Matrix Indicator (TPMI) indicating the codebook index used from the codebook for MIMO.

3. *Data Transfer:* The eNodeB sends user data on PDSCH.

4. *Link Adaptation:* Signal transmission parameters are changed with time and frequency in order to better adjust to the changing radio link conditions. The set of techniques used for this purpose are known as link adaptation techniques.
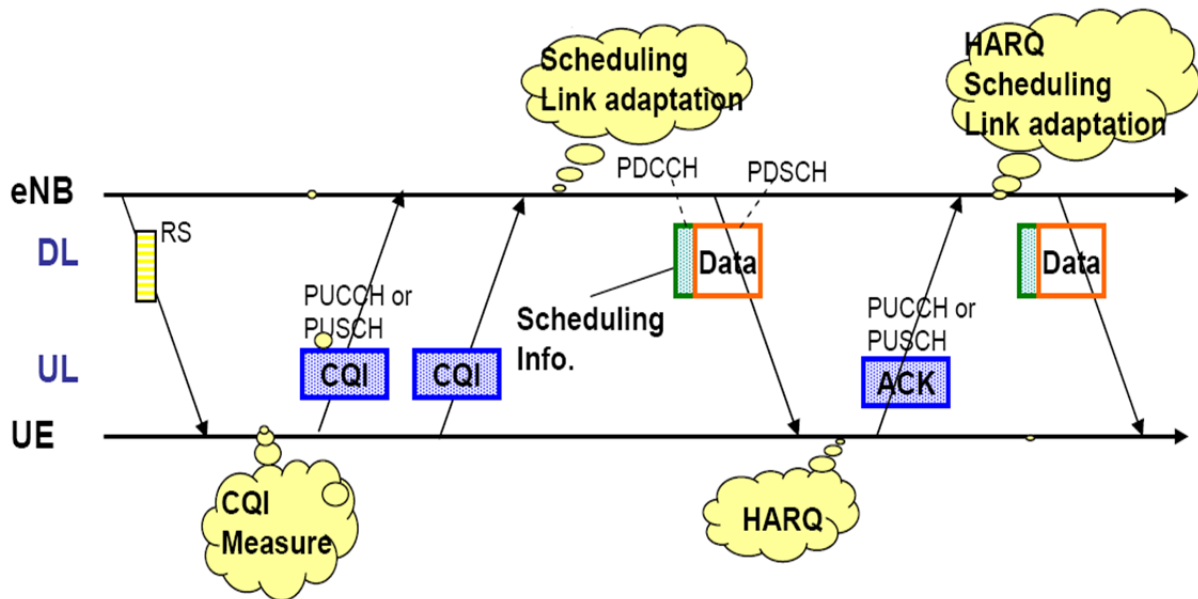
Figure 3.7: Use of physical channels during downlink data transfer

5. *HARQ ACK/NACK:* The UE attempts to decode the received packets. The UE sends HARQ ACK or NACK based on its success or failure to decode the packets respectively. The HARQ ACK/NACK is sent on PUCCH if there is no uplink data transfer and on PUSCH if there is ongoing uplink data transfer.

The data transfer in uplink involves a few basic steps as follows and it uses the physical channels shown in Section 2.5.4. Fig. 3.8 explains this transfer.

1. *Seeking Resource Allocation:* When the UE has data available for transmission and it needs uplink resources allocation, it notifies the eNodeB seeking resources. For this purpose, the UE sends Buffer Status Reports (BSRs) on PUSCH if there is no ongoing data transfer on PUSCH and it sends Scheduling Request (SR) on PUCCH if there is no ongoing data transfer over PUSCH.

2. *Allocation of Resources:* The eNodeB informs the UE about the allocated resource blocks on PUSCH and configuration for data transmission using PDCCH. The UE decodes PDCCH and obtains necessary configuration for the allocated resources. The configuration includes Modulation and Coding Scheme (MCS) indicating transport format.
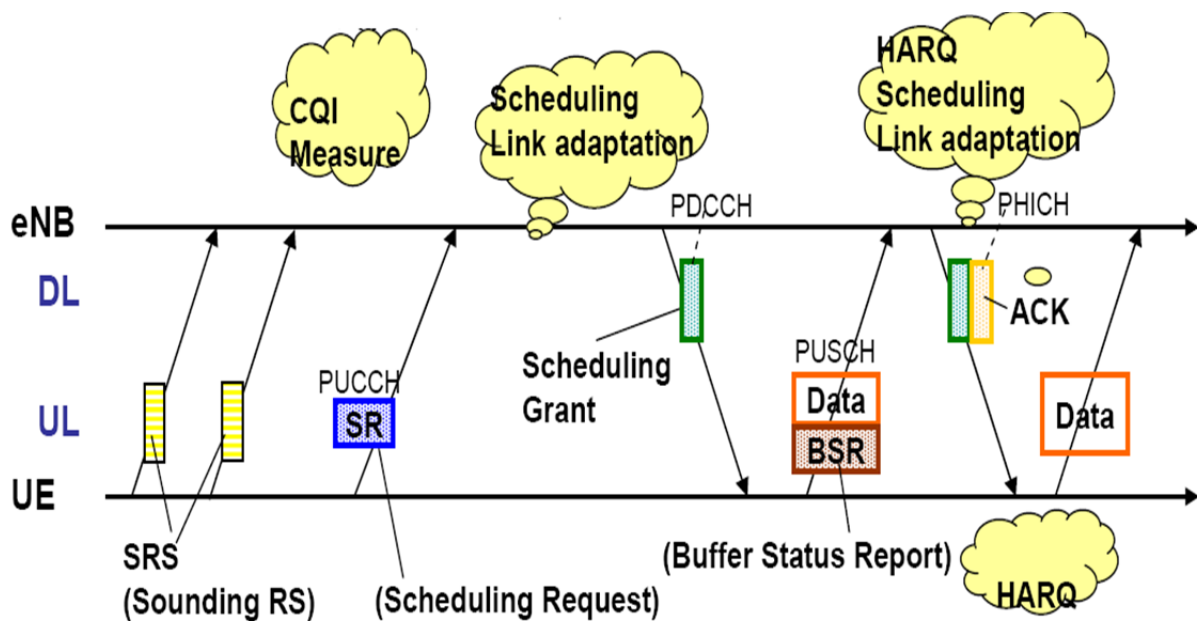
Figure 3.8: Use of physical channels during uplink data transfer

3. *Data Transfer:* The UE sends user data over the allocated resources on PUSCH.
4. *Link Adaptation:* Link adaptation techniques are applied to adapt the signal transmission parameters to the changing radio link conditions.
5. *HARQ ACK/NACK:* The eNodeB attempts to decode the received packets and sends HARQ ACK/NACK on PHICH indicating success or failure in decoding.

3.7 Conclusion

LTE has only eNodeB in its radio access network and therefore, it is said to have a flat architecture. The access strata (AS) functioning between radio network and user equipment over this flat architecture, have diversified features. The functions are split among different layers and sublayers. The layers and sublayers of the radio network and the user equipment have peer-to-peer communication to conduct the functions. This chapter gives a brief overview of the functionalities of AS.

The technology of power storage or battery life has not sufficiently improved to cope with the advance in other technologies that run on batteries. Thus, LTE, which is introducing many different applications, encounter difficulties with the battery power consumption of the UE. Therefore, there are efforts to improve the battery power saving of the UE. It can be possible to modify the control of AS in order to extend the battery life and such a modification will be discussed in the next chapter.

Chapter 4

MAC/RRC Control over DRX to Improve Power Saving for NRT Traffic

The incredibly higher data rates and better quality of services have motivated the advent and use of many different applications in LTE. This causes the battery of the user equipment (UE) to drain out quickly and the battery power saving has become a significant concern. The key technique to save power and prolong battery life is to let the UE switch off the receiver circuitry periodically. This is referred to as discontinuous reception (DRX). An improvement in DRX power saving usually leads to a potential increase in the packet delay. This chapter considers the possibilities of improvement in DRX power saving while exploiting permissible increase in packet delay.

4.1   Introduction

The discontinuous reception (DRX) operation is included in LTE to achieve power saving and prolonged battery life of the UE. The DRX operation was explained in Section 3.5.1. The MAC layer performs DRX in the way it is configured by the RRC layer. The DRX cycle length involves a trade-off between power saving and packet delay. When a packet arrives during the off duration of DRX cycle, it cannot be scheduled until the off duration is over. This enhances the packet delay, which is the main drawback of DRX operation. A longer DRX cycle allows better power saving but with a longer packet delay.

An optimum DRX configuration depends on the current data traffic, which is not easy to estimate accurately, especially, for nonreal-time applications. In this paper, we propose a variation in the DRX cycle length avoiding continuous estimation of the data traffic when only nonreal-time applications are running with no real-time applications active. Since small delay in nonreal-time traffic does not essentially impact the user's experience adversely, we allow limited amount of delay deliberately in our proposal to attain significant improvement in power saving. Our proposal also improves the delay in service resumption after a long inactivity. We use a stochastic analysis to validate the improvement of the proposal.

4.2 States Associated with DRX

This chapter considers DRX in RRC_CONNECTED state only. The various states associated with DRX are shown in Fig. 4.1. State A indicates ordinary data transfer and T_D^{AO} represents its duration. State B indicates running inactivity timer at the end of ordinary data transfer and T_I^{BD} represents its duration. In state B, the inactivity timer is reset before it expires because of the arrival of a packet. State C indicates data transfer after state B and T_D^{AI} represents the duration of state C. After every state C, the inactivity timer restarts. N_I^D represents how many times the inactivity timer restarts.
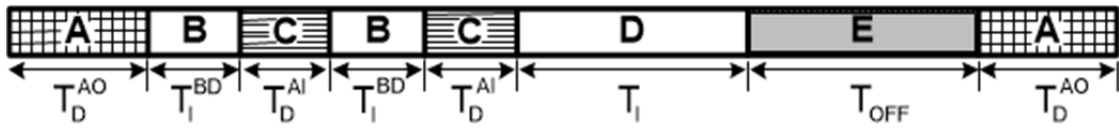
Figure 4.1: States associated with DRX

State D indicates running inactivity timer when the timer can run up to its expiry with no packet arrivals. T_I represents the whole duration of the inactivity timer. State E indicates running DRX cycles after the expiry of the inactivity timer. The short cycles would run first and there can be N number of short cycles at the maximum. After N number of short cycles, long cycles start running. T_S and T_L represent the length of short and long DRX cycles, respectively. T_{OFF} represents the duration of state E. DRX operation is terminated if new packets arrive and then ordinary data transfer starts. So, state A appears after state E. The notations and states, shown in this section, will be used in the rest of this chapter.

4.3 Related Work

[26]−[29]investigated the impact of DRX parameter values on power saving and delay. Such investigation was also performed for the industrial DRX model proposed by Nokia [30]. It is shown that a good power saving can be achieved at the cost of delay. Considering this fact, [31] proposes an algorithm to find feasible ranges of DRX parameter values for maximizing power saving while satisfying a specified delay constraint as well as minimizing delay while satisfying a specified power saving. [32] proposes a scheme to select DRX configuration by jointly formulating power saving and delay and considering the operator's preference of power saving over delay. [33] proposes scaling the length of both short and long DRX cycles to choose a desired balance between power saving and delay. [34]

proposes varying on duration in addition to the length of short and long cycles. [34] also proposes switching between short and long cycles using various patterns for desired performance. [35] proposes an algorithm to select appropriate DRX configuration while keeping the delay below a threshold. While many proposals are made to reach an appropriate DRX configuration, the current data traffic largely dictates which configuration may be appropriate and so, adaptation of DRX parameters based on the current traffic can be more efficient. [27],[28],[35] show that continuous adaptation of DRX parameters can obtain better performance compared to fixed DRX parameters. [36],[37] also suggest adaptation based on the current traffic. For estimation of the current traffic, [36] uses a single bit feedback from the UE known as power preference indication (PPI), which has been adopted in 3GPP Release 11. [37] suggests a method of quick estimation of the traffic using the maximum likelihood method and the properties of Gamma distribution. [38] proposes adaptation based on the traffic where the time instants to reconfigure DRX parameters also depend on the traffic and shows an enhancement of power saving considering the delay constraint. [39] proposes the use of a fuzzy logic controller (FLC) to adjust DRX cycle after the traffic is estimated.

[40] proposes an algorithm to adjust DRX parameters when the delay crosses predefined high and low threshold values. [40] also proposes a second algorithm, which additionally considers the recently used Modulation and Coding Scheme (MCS) to cross predefined threshold values. However, all these threshold values need to be adaptive for an optimum performance in practice and proper selection or adaptation of the threshold values may not be easy. [41] proposes that the UE sends an indication to the eNodeB when the application is delay tolerant. Then the eNodeB can use DRX MAC control element to command the UE to switch to long DRX cycles from short DRX cycles. However, no sides may be fully aware of the potential delay in packet arrival and so, the decision for switching to long cycle may not be very optimal.

4.4   Difference in Delay between RT and NRT Services

The quality-of-service (QoS) requirements are quite different between real-time (RT) and nonreal-time (NRT) traffic and so, they assume different QCI as shown in Table 2.3. The examples of NRT applications are buffered streaming video, web browsing, e-mail, chat, FTP, and P2P file sharing,. On the other hand, the examples of RT applications are voice, live

streaming video and online gaming. RT traffic has a stringent delay requirement, while some level of packet loss rate can be acceptable. Conversely, NRT traffic is error-sensitive but the transmission delay can be larger. The PELR is typically 10^{-3} and 10^{-6} for RT and NRT traffic, respectively. The PDB is typically 100 ms and 300 ms for RT and NRT traffic, respectively [42]. The processing in AS depends on QCI of the EPS bearer and consequently, RT and NRT services differ in their processing delays. The scheduling of radio resources at the MAC layer uses algorithms, which treat RT and NRT traffic differently [43]. Similarly, the RLC layer uses different modes for RT and NRT traffic. The RT bearers typically use unacknowledged mode (UM) of RLC since they are error-tolerant. Conversely, NRT applications typically use acknowledged mode (AM) of RLC, which perform retransmissions in order to limit the errors.

The major factor that can be attributed to the difference in the processing delay between RT and NRT traffic is the use of RLC modes. This is because the retransmissions in RLC AM incur some delay as explained in Section 3.4.3. The overall delay in the correction of data can be estimated as the summation of the reordering timer value and the round trip time (RTT). The RTT includes round trip propagation delay and processing delay at both ends. The reordering timer value is configured by the eNodeB as was shown in Section 3.4.3.

Let us assume that $E[D_{RT}]$ and $E[D_{NRT}]$ represent the expected delay for RT and NRT traffic, respectively. $E[D_{RTX}]$ represents the expected delay for RLC AM retransmissions, which occurs only in the case of NRT traffic. Ignoring all minor causes of differences in the delays between RT and NRT traffic, in crude terms, these delays can be related as

$$E[D_{NRT}] = E[D_{RT}] + E[D_{RTX}]. \tag{4.1}$$

The PDB for NRT traffic is much larger than that of RT traffic because of the delay tolerance of NRT applications. Thus, $PDB_{NRT} - PDB_{RT} > E[D_{NRT}] - E[D_{RT}]$, where PDB_{NRT} and PDB_{RT} denote the PDB for NRT and RT traffic, respectively. This gives

$$PDB_{NRT} - PDB_{RT} = E[D_{NRT}] - E[D_{RT}] + \bar{\zeta} \tag{4.2}$$

where $\bar{\zeta}$ can be expressed using (1) as

$$\bar{\zeta} = PDB_{NRT} - PDB_{RT} - E[D_{RTX}]. \tag{4.3}$$

4.5 Proposed DRX Scheme

A continuous increase in the DRX cycle length followed by a continuous decrease in the length is proposed [2]. The RRC layer configures the parameters in the proposed variation using DRX-Config IE. The proposal can be applied only when nonreal-time (NRT) applications are running with no real-time (RT) applications active. The proposal improves the power saving while increasing the delay of NRT traffic within a permissible limit. Unlike RT traffic, the NRT traffic permits the delay requirement to slacken to some extent. In other words, a limited amount of delay can be allowed to NRT traffic essentially with no adverse impact on the user's experience. Therefore, the allowable delay for NRT traffic is exploited to improve power saving in the proposal.

The processing in the access strata (AS) of the protocol stack depends on the QoS class identifier (QCI) of the data traffic. QCI takes on various attributes for different types of applications. Thus, RT and NRT traffic are treated differently and consequently, NRT traffic takes longer processing delay compared to RT traffic. This longer processing delay consumes part of the additional delay, permissible for NRT traffic in comparison with RT traffic. Nevertheless, after this consumption, it can be shown that some additional delay is left over. The proposal exploits this left over delay and so, it increases the delay of NRT traffic within a permissible limit deliberately.

The proposal avoids the continuous estimation of traffic. In fact, the estimation of current traffic, used in many proposals as mentioned earlier, increases complicacy and burden. More importantly, the NRT traffic is highly irregular and unpredictable. So, the traffic estimation for NRT traffic cannot be very accurate and an adaptation of DRX configuration based on this estimation will not result in a very optimum outcome in practice. The proposal functions for delay tolerant applications, as considered in [40] and [41], but avoiding the overhead of the messages and admitting the uncertainty in packet arrivals.

The proposal additionally reduces the delay in service resumption when there has been a long inactivity without terminating an NRT session. Such long inactivity may result from various reasons. For example, the server application may get busy and may not respond for a while. Also, the user may take a break to take tea, coffee or lunch, to discuss with co-workers, to use the restroom, etc. In such cases, the DRX runs for a long period over the inactivity. In the existing scheme, the data transfer takes quite a while to resume after the inactivity and this can be a nuisance.

The proposed scheme is based on the fact that $PDB_{NRT} - PDB_{RT} = E[D_{NRT}] - E[D_{RT}]$ can be permissible in the system. Therefore, according to (2), $E[D_{NRT}]$ can be allowed to be increased to a new value $E[D_{NRT}]_{new}$ given by

$$E[D_{NRT}]_{new} = E[D_{NRT}] + \Delta_{NRT} \qquad (4.4)$$

where $\Delta_{NRT} \leq \bar{\zeta}$. $\qquad (4.5)$

Here, (4.5) sets a constraint. In the proposal, $E[D_{NRT}]$ is increased purposely to achieve improvement in power saving while satisfying the constraint in (4.5). To accomplish it in a simple method, the DRX cycle length is proposed to start increasing linearly with time. A simple method can invoke small changes in the associated protocols and help easy and quick implementation. In the proposal, instead of allowing a few short cycles in the beginning, the DRX cycle length keeps increasing from T_S by a fixed step size T_{st} until it reaches T_L. Let us assume that n number of cycles are taken to reach T_L. Once n DRX cycles have been executed with no packet arrivals, it can be assumed that there has been already a long inactivity in the session. The proposal attempts to limit delay in service resumption after a long inactivity. Since a smaller DRX cycle length reduces this delay, after n DRX cycles, the DRX cycle length starts decreasing from T_L to T_S by the same fixed step size T_{st}. After completion of this decline up to T_S, assuming that the service may not resume soon, a rise in the DRX cycle length with previous pattern is again used. The rise will be followed by a fall like before for the same reason. These alternate repetitions continue leading to a triangular fashion of variation in the DRX cycle length as illustrated in Fig. 4.2. The smooth gradual variation in the DRX cycle length can be expected to match a change in the data traffic to some extent and thus, the estimation of data traffic is avoided. The RRC layer at eNodeB configures T_S, T_L and T_{st} using DRX-Config IE when a session is set up and for this configuration, the eNodeB considers the operator's choice and an overall estimate of the potential traffic. The period of the triangular variation is $T_T = 2nT_s + n(n + 1)T_{st}$. The numerical examples in Section 4.7 show that the proposed scheme can satisfy the constraint in (4.5).
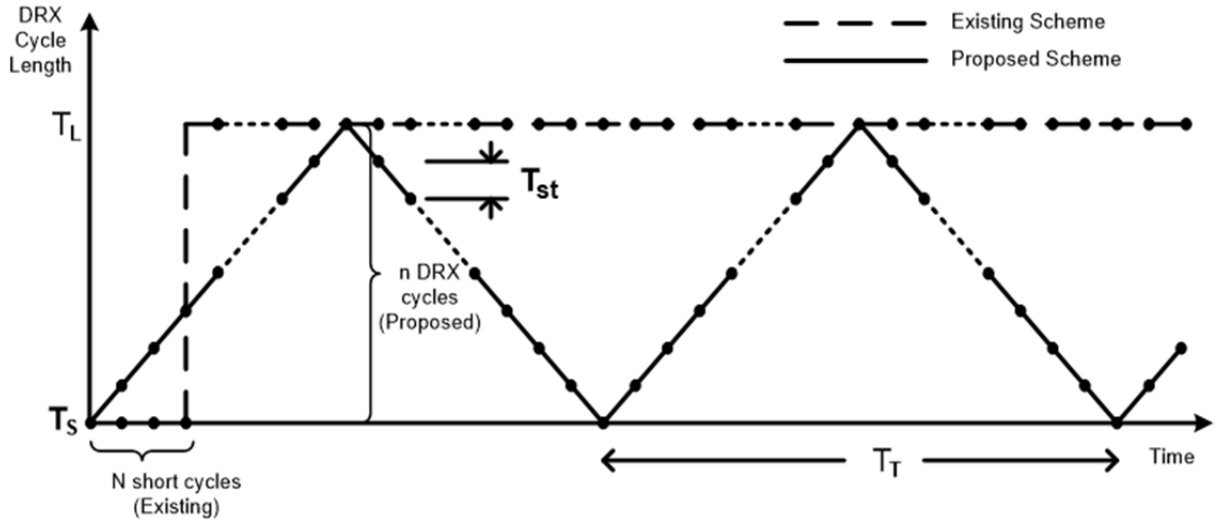
Figure 4.2: Existing and proposed schemes for DRX cycles

4.5.1   *Estimation of Power Saving in DRX*

A new factor is proposed to estimate the power saving in DRX analysis. Various factors have been used so far. [44] uses percentage power saving as

$$\left(\frac{E[T_{Data}]}{E[T_{Total}]} + \frac{P_{OFF}E[T_{OFF}]}{P_{Data}E[T_{Total}]} \right) \times 100$$

where the power consumption during data transfer and during the period of DRX cycles are denoted as P_{Data} and P_{OFF}, respectively. The expected durations of these two states are denoted as $E[T_{Data}]$ and $E[T_{OFF}]$, respectively. $E[T_{Total}]$ represents the mean whole duration including data transfer and DRX. [28] uses power saving factor as

$$\frac{E[T_S - T_{ON}] + E[T_L - T_{ON}]}{E[T_{Total}]}$$

where $E[T_S - T_{ON}]$ and $E[T_L - T_{ON}]$ represent the expected short and long DRX periods, respectively, excluding the on durations. [45], [46] use power saving factor as $\frac{E[T_{OFF}]}{E[T_{Total}]}$ without incorporating any power levels. However, there is a significant difference among the power consumptions during effective data transfer, during the period while the inactivity timer runs looking for data and during DRX. So, the incorporation of power levels in power saving estimation is suggested but introducing a third power level P_{wait} denoted as the power consumption while the inactivity timer runs. Secondly, the power consumption becomes extremely low for DRX cycles of any length [47]-[49]. In fact, the impact of differences in power consumption for DRX cycles of different lengths and for on durations of different

66

lengths is trivial and so, such differences are ignored for simplicity. The average power consumption is defined as

$$P_{avg} = \frac{P_{Data} \cdot E[T_{Data}] + P_{wait} \cdot E[T_{wait}] + P_{OFF} \cdot E[T_{OFF}]}{E[T_{Total}]} \tag{4.6}$$

where $E[T_{wait}]$ denotes the expected duration of running the inactivity timer and

$$E[T_{Total}] = E[T_{Data}] + E[T_{wait}] + E[T_{OFF}]. \tag{4.7}$$

To better indicate the practical power saving in DRX, the proposed percentage power consumption is defined as

$$\frac{P_{avg}}{P_{Data}} = \left(\frac{E[T_{Data}]}{E[T_{Total}]} + \frac{P_{wait}E[T_{wait}]}{P_{Data}E[T_{Total}]} + \frac{P_{OFF}E[T_{OFF}]}{P_{Data}E[T_{Total}]} \right) \times 100. \tag{4.8}$$

4.6   Analytical Model for Proposed and Existing Schemes

Analytical models have been derived in this section in order to evaluate the proposed and existing schemes.

4.6.1   Queuing System Consideration

The queuing systems, in general, may be characterized by complex input process, service time distribution, number of servers, buffer size and queue disciplines. A commonly used shorthand notation, called Kendall notation, for the queue model describes the arrival process, service distribution, the number of servers and the buffer size as follows:
{arrival process}/{service distribution}/{number of servers}/{buffer size}-{queue discipline}

The commonly used characters for the first two positions in this shorthand notation are as follows.

- – D (Deterministic)
- – M (Markovian - Poisson for the arrival process or exponential for the service time distribution)
- – G (General)
- – GI (General and independent)
- – Geom (Geometric)

The fourth position is used for the number of buffer places including the buffer spaces available at the servers. The fourth position is not used if the waiting room is unlimited. The

fifth position is used for the queue discipline. Thus, M/G/1 denotes a single-server queue (SSQ) with Poisson arrivals and generally distributed service time requirement.

The analytical models for the proposed and existing schemes assume M/G/1 system queue. So, the packet arrival follows Poisson process and the inter-arrival times are distributed exponentially. The Poisson process is a continuous-time point process which is memory less and orderly. It applies to many cases where a certain event occurs at different points in time. Every point in time has the same chance of having an occurrence, i.e. occurrences are equally likely to happen at all times. This property is also called time-homogeneity. [50].

It is assumed that the packet arrival rate is λ and the service or transmission rate of the packets is μ. The mean transmission time of a packet is $E[S] = 1/\mu$. The traffic intensity can be expressed as $\rho = \lambda/\mu$. The system is considered stable with $\mu > \lambda$. The probability of packet arrival in t_1 duration is $\int_0^{t_1} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t_1}$. Thus, the probability of no packet arrival in t_1 duration is $1 - \int_0^{t_1} \lambda e^{-\lambda t} dt = e^{-\lambda t_1}$.

The percentage power consumption is computed using (4.8) and for this purpose, $E[T_{OFF}]$, $E[T_{wait}]$ and $E[T_{Data}]$ are evaluated below for the proposed and existing schemes. Since, the Poisson traffic model is short range dependent, the early events will be prominent in the analysis.

4.6.2   Mean Duration of DRX Operation

The expected duration of continuous DRX operation in the proposed scheme $E[T_{OFF}]_{prop}$ can be given by

$$E[T_{OFF}]_{prop} = E[T_R] + E[T_F] \tag{4.9}$$

where $E[T_R]$ and $E[T_F]$ represent the mean duration of DRX operation in the rise parts and the fall parts of the triangular periods, respectively. The probability that no packet arrives in j number of triangular periods and (k–1) number of DRX cycles of the rise part of (j+1) numbered triangular period and thereafter packet arrives in the next DRX cycle (that is, the kth cycle of the rise part), can be expressed as

$$\chi_{j,k}^R = e^{-j\lambda T_T} R_k . \tag{4.10}$$

Here, R_k is given by

$$R_k = e^{-\delta_{k,R}\lambda} \left[1 - e^{-\lambda(T_s + kT_{st})} \right] \tag{4.11}$$

68

where $\delta_{k,R}$ is given by

$$\delta_{k,R} = (T_s + T_{st}) + (T_s + 2T_{st}) + \cdots\cdots + [T_s + (k-1)T_{st}]$$

$$= (k-1)T_s + \frac{k(k-1)T_{st}}{2}. \tag{4.12}$$

The whole duration of DRX associated with $\chi_{j,k}^R$ is given by

$$T_{j,k}^R = j\,T_T + (T_s + T_{st}) + (T_s + 2T_{st}) + \cdots\cdots + (T_s + kT_{st})$$

$$= jT_T + kT_s + \frac{k(k+1)T_{st}}{2} \tag{4.13}$$

$E[T_R]$ can be expressed as

$$E[T_R] = \sum_{j=0}^{\infty}\sum_{k=1}^{n} \chi_{j,k}^R \cdot T_{j,k}^R$$

$$= \sum_{k=1}^{n} R_k\left[T_T \sum_{j=0}^{\infty} je^{-j\lambda T_T} + \left\{ kT_s + \frac{k(k+1)T_{st}}{2} \right\} \sum_{j=0}^{\infty} e^{-j\lambda T_T} \right]$$

$$= \frac{e^{-\lambda T_T}}{(1-e^{-\lambda T_T})^2}\left[T_T \sum_{k=1}^{n} R_k + (1 - e^{-\lambda T_T}) \sum_{k=1}^{n} R_k\left\{ kT_s + \frac{k(k+1)T_{st}}{2} \right\} \right]. \tag{4.14}$$

Similarly, the probability that no packet arrives in j number of triangular periods and in the $T_T/2$ long rise part and (k−1) number of DRX cycles of the fall part of (j+1) numbered triangular period and thereafter packet arrives in the next DRX cycle (that is, the kth cycle of the fall part), can be expressed as

$$\chi_{j,k}^F = e^{-j\lambda T_T}e^{-\lambda(T_T/2)}F_k. \tag{4.15}$$

Here, F_k is given by

$$F_k = e^{-\delta_{k,F}\lambda}\left[1 - e^{\lambda\{T_s + (n-k)T_{st}\}}\right] \tag{4.16}$$

where $\delta_{k,F}$ is given by

$$\delta_{k,F} = \{T_s + (n-1)T_{st}\} + \cdots + [T_s + \{n - (k-1)\}T_{st}]$$

$$= (k-1)T_S + \frac{(2n-k)(k-1)T_{st}}{2}. \tag{4.17}$$

The whole duration of DRX associated with $\chi_{j,k}^F$ is given by

$$T_{j,k}^F = jT_T + \frac{T_T}{2} + [T_s + (n-1)T_{st}] + .. + [T_s + (n-k)T_{st}]$$

$$= jT_T + \frac{T_T}{2} + kT_s + T_{st}\left[nk - \frac{k(k+1)}{2}\right]. \tag{4.18}$$

69

$E[T_F]$ can be expressed as

$$E[T_F] = \sum_{j=0}^{\infty} \sum_{k=1}^{n} \chi_{j,k}^F \cdot T_{j,k}^F$$

$$= e^{-\lambda(T_T/2)} \sum_{k=1}^{n} F_k \left[T_T \sum_{j=0}^{\infty} j e^{-j\lambda T_T} + \left[\frac{T_T}{2} + kT_s + T_{st} \left\{ nk - \frac{k(k+1)}{2} \right\} \right] \sum_{j=0}^{\infty} e^{-j\lambda T_T} \right]$$

$$= \frac{e^{-\frac{3\lambda T_T}{2}}}{(1-e^{-\lambda T_T})^2} \left[T_T \sum_{k=1}^{n} F_k + (1 - e^{-\lambda T_T}) \sum_{k=1}^{n} F_k \left[\frac{T_T}{2} + kT_s + T_{st} \left\{ nk - \frac{k(k+1)}{2} \right\} \right] \right].$$

(4.19)

$E[T_{OFF}]_{prop}$ can now be determined using (4.9), (4.14) and (4.19). The mean duration of DRX in the existing scheme $E[T_{OFF}]_{exist}$ is given by

$$E[T_{OFF}]_{exist} = E[T_S] + E[T_L]$$

(4.20)

where $E[T_S]$ and $E[T_L]$ represent the mean duration of short and long cycles, respectively.

The probability that (i–1) number of short cycles occur with no packet arrival and thereafter packet arrives in i^{th} short cycle, can be expressed as

$$\chi_i^S = (e^{-\lambda T_s})^{i-1} (1 - e^{-\lambda T_s}).$$

(4.21)

The mean number of short cycles $E[N_{SC}]$ is given by

$$E[N_{SC}] = \sum_{i=1}^{N} i \cdot \chi_i^S = e^{\lambda T_s} (1 - e^{-\lambda T_s}) \sum_{i=1}^{N} i (e^{-\lambda T_s})^i$$

$$= \frac{(1-e^{-N\lambda T_s})}{(1-e^{-\lambda T_s})} - N e^{-N\lambda T_s}.$$

(4.22)

$E[T_S]$ is given by

$$E[T_S] = T_s E[N_{SC}] = T_S \left[\frac{(1-e^{-N\lambda T_s})}{(1-e^{-\lambda T_s})} - N e^{-N\lambda T_s} \right].$$

(4.23)

The probability that no packet arrives in N number of short cycles and (j–1) number of long cycles and thereafter, packet arrives in j^{th} long cycle, can be expressed as

$$\chi_j^L = e^{-N\lambda T_s} \left(e^{-\lambda T_L} \right)^{j-1} \left(1 - e^{-\lambda T_L} \right).$$

(4.24)

So, the mean number of long cycles $E[N_{LC}]$ is given by

$$E[N_{LC}] = \sum_{j=1}^{\infty} j\chi_j^L = e^{-N\lambda T_S} \left(1 - e^{-\lambda T_L}\right) \sum_{j=1}^{\infty} j\left(e^{-\lambda T_L}\right)^{j-1}$$

$$= \frac{e^{-N\lambda T_S}}{1 - e^{-\lambda T_L}}. \tag{4.25}$$

$E[T_L]$ is given by

$$E[T_L] = T_L E[N_{LC}] = \frac{T_L e^{-N\lambda T_S}}{1 - e^{-\lambda T_L}}. $$

$$\tag{4.26}$$

$E[T_{OFF}]_{exist}$ can now be determined using (4.20), (4.23) and (4.26).

4.6.3   Mean Duration of Running Inactivity Timer

The probability P_x that the inactivity timer restarts x times because of packet arrivals and thereafter, the inactivity timer expirers once completing T_I, can be expressed as

$$P_x = \left(1 - e^{-\lambda T_I}\right)^x e^{-\lambda T_I}. \tag{4.27}$$

So, the mean number of restart of the inactivity timer $E[N_I^D]$ is given by

$$E[N_I^D] = \sum_{x=0}^{\infty} xP_x$$

$$= e^{-\lambda T_I} \sum_{x=0}^{\infty} x\left(1 - e^{-\lambda T_I}\right)^x = e^{\lambda T_I} - 1. \tag{4.28}$$

As the inactivity timer runs, the occurrences in the scenario are limited to the expiry of the timer, which means that the occurrences lie only within the range as long as the timer runs. In statistics, a truncated distribution is a conditional distribution that results from restricting the domain of some other probability distribution. Truncated distributions arise in in cases where the ability to record, or even to know about, occurrences is limited to values which lie above or below a given threshold or within a specified range. Thus, the truncated exponential distribution can be used for the packet arrivals to estimate T_I^{BD}. The truncated exponential distribution can be expressed as

$$f(t) = \frac{\lambda\, e^{-\lambda t}}{1 - e^{-\lambda T_I}}, \quad 0 \le t \le T_I$$

$$= 0. \qquad\qquad t > T_I \tag{4.29}$$

$E[T_I^{BD}]$ can be expressed as

$$E[T_I^{BD}] = \int_0^{\infty} tf(t)dt = \frac{\lambda}{1 - e^{-\lambda T_I}} \int_0^{T_I} te^{-\lambda t} dt$$

$$= \frac{1}{\lambda} - \frac{T_I}{e^{\lambda T_I} - 1}. \tag{4.30}$$

The mean duration of running inactivity timer is the summation of the average period in state B and state D. Since the average period in state B is $E[T_I^{BD}]E[N_I^D]$ and the period in state D is T_I, $E[T_{wait}]$ for both the existing and the proposed methods can be expressed as

$$E[T_{wait}] = E[T_I^{BD}]E[N_I^D] + T_I$$
$$= \frac{e^{\lambda T_I - 1}}{\lambda}. \tag{4.31}$$

4.6.4 Mean Duration of Data Transfer

The mean duration of data transfer is the summation of the average period in state A and state C. Since $E[N_I^D]$ is the mean number of restart of the inactivity timer, the mean duration of data transfer $E[T_{Data}]$ for both the existing and the proposed methods can be given by

$$E[T_{Data}] = E[T_D^{AO}] + E[T_D^{AI}]E[N_I^D]. \tag{4.32}$$

When a single packet is transmitted, its mean transmission time is $E[S]=1/\mu$. The probability of m number of packet arrivals in this $1/\mu$ duration P_m is given by

$$P_m = \frac{e^{-\lambda/\mu}\left(\lambda/\mu\right)^m}{m!} = \frac{e^{-\rho} \cdot \rho^m}{m!}. \tag{4.33}$$

So, the mean number of packet arrivals in $1/\mu$ duration can be shown as

$$\sum_{m=0}^{\infty} mP_m = \sum_{m=0}^{\infty} m \frac{e^{-\rho} \cdot \rho^m}{m!} = \rho. \tag{4.34}$$

If one packet arrives while the inactivity timer is running, then within its transmission period, additional ρ number of packets arrive on average. These additional packets require additional ρ/μ transmission period in which further $\rho\lambda/\mu = \rho^2$ number of packets arrive on average. This ideally grows up and the total expected transmission period is given by

$$E[T_D^{AI}] = \frac{1}{\mu} + \frac{\rho}{\mu} + \frac{\rho^2}{\mu} + \frac{\rho^3}{\mu} + \cdots\cdots$$
$$= \frac{1}{\mu} \sum_{k=0}^{\infty} \rho^k = \frac{1}{\mu(1-\rho)}. \tag{4.35}$$

$E[T_D^{AO}]$ can be determined similarly. During the DRX period $E[T_{OFF}]$, $\lambda E[T_{OFF}]$ number of packets arrive. These packets require $\lambda E[T_{OFF}]/\mu = \rho E[T_{OFF}]$ transmission period. In this

period, additional $\lambda \rho E[T_{OFF}]$ number of packets arrive, which require additional $\lambda \rho E[T_{OFF}]\big/_\mu = \rho^2 E[T_{OFF}]$ transmission period. Thus, $E[T_D^{AO}]$ is given by

$$E[T_D^{AO}] = E[T_{OFF}]\{\rho + \rho^2 + \cdots\cdots\cdots\}$$
$$= E[T_{OFF}]\sum_{k=1}^{\infty}\rho^k = \frac{\rho}{1-\rho}\,E[T_{OFF}]. \tag{4.36}$$

Using equations (28), (32), (35) and (36), $E[T_{Data}]$ can be expressed as

$$E[T_{Data}] = \frac{\rho}{1-\rho}\,E[T_{OFF}] + \left(e^{\lambda T_I} - 1\right)\frac{1}{\mu(1-\rho)}$$
$$= \frac{\rho}{1-\rho}\left\{E[T_{OFF}] + \frac{e^{\lambda T_I} - 1}{\lambda}\right\}$$
$$= \frac{\rho}{1-\rho}\{\,E[T_{OFF}] + E[T_{wait}]\} \tag{4.37}$$

$E[T_{Data}]_{exist}$ and $E[T_{Data}]_{prop}$ can be evaluated for the existing and the proposed methods, respectively, using (4.9), (4.20), (4.31) and (4.37). Similarly, $E[T_{Total}]_{exist}$ and $E[T_{Total}]_{prop}$ can be evaluated for the existing and the proposed methods, respectively, using (4.7).

4.6.5   *Delay Performance*

The delay analysis will be different between the cases when the UE is running DRX and when the UE is not running. Let us assume that N_Q represents the number of packets in the queue but excluding those for whom transmission is ongoing and W_Q represents the waiting time of a packet from the moment it arrives until its transmission commences. According to Little's formula, the mean number of customers in the queue in steady-state excluding the customer in service $E[N_Q]$ and the mean delay of a customer, in steady-state, from the moment it arrives until its service commences (waiting time in the queue) can be $E[W_Q]$ related as follows.

$$E[N_Q] = \lambda\,E[W_Q]. \tag{4.38}$$

The mean residual service time $E[R]$ represents the mean service or transmission time of packets currently in transmission when a packet arrives. According to Pollaczek

73

Khintchine formula, the waiting time in the queue of an arriving packet to an M/G/1 queue is the remaining service time of the packet in service plus the sum of the service times of all the packets in the queue ahead of the arriving packet. Since $\frac{E[N_Q]}{\mu}$ represents the service time of all packets ahead in the queue waiting for service, the mean wait time in the queue is given by

$$E[W_Q] = E[R] + \frac{E[N_Q]}{\mu}. \tag{4.39}$$

Using Little's formula,

$$E[W_Q] = E[R] + \rho E[W_Q]$$

$$= \frac{E[R]}{1-\rho}. \tag{4.40}$$

When the transmission rate is very high compared to the packet arrival rate, making $\mu \gg \lambda$, the packets ahead in the queue can be quickly transmitted resulting in $E[W_Q] \approx E[R]$. $E[R]$ has the relationship [46]

$$E[R] = \frac{\lambda E[S^2]}{2}. \tag{4.41}$$

Using (4.40),

$$E[W_Q] = \frac{\lambda E[S^2]}{2(1-\rho)}. \tag{4.42}$$

While the UE is not in DRX (in states A, B, C or D), the packet delay can be given by

$$E[D_{N\_DRX}] = E[W_Q] = \frac{\lambda E[S^2]}{2(1-\rho)}. \tag{4.43}$$

On the other hand, while the UE is running DRX (in state E), the packet may arrive anytime within the DRX cycle. Then there will be an additional delay $E[W_D]$ because the packets are not processed until the particular DRX cycle is over. The overall mean delay $E[D_{DRX}]$ can be given by

$$E[D_{DRX}] = E[W_Q] + E[W_D]. \tag{4.44}$$

The Poisson process has time-homogeneity as mentioned in Section 4.6.1. Thus, the packet can arrive at any time during the DRX cycle with equal probability and it has to wait for the rest of the DRX cycle for any process to begin. The mean wait time can be computed as the half of the average length of the DRX cycle and it is independent of the packet arrival

rate. This mean wait time is equivalent to $E[R]$ and $E[R]_{prop}$ for either the rise time or the fall time of the proposed method can be expressed as

$$E[R]_{prop} = \frac{1}{n}\left[\frac{T_s + T_{st}}{2} + \frac{T_s + 2T_{st}}{2} + .. + \frac{T_s + nT_{st}}{2}\right]$$

$$= \frac{2T_s + (n+1)T_{st}}{4}. \tag{4.45}$$

Using (4.40) and (4.45), $E[W_D]_{prop}$ can be expressed as

$$E[W_D]_{prop} = \frac{2T_s + (n+1)T_{st}}{4(1-\rho)}. \tag{4.46}$$

The mean delay in the case of rise time $E[D_{DRX}]_R$ and fall time $E[D_{DRX}]_F$ can be expressed using (4.42), (4.44) and (4.46) as

$$E[D_{DRX}]_R = E[D_{DRX}]_F = E[W_Q] + E[W_D]_{prop}$$

$$= \frac{\lambda E[S^2]}{2(1-\rho)} + \frac{2T_s + (n+1)T_{st}}{4(1-\rho)}. \tag{4.47}$$

Denoting the probability of packet arrival during the rise time and the fall time of the proposed method as γ_R and γ_F, respectively, they can be expressed as

$$\gamma_R = \frac{E[T_R]}{E[T_{Total}]_{prop}} \tag{4.48}$$

and

$$\gamma_F = \frac{E[T_F]}{E[T_{Total}]_{prop}}. \tag{4.49}$$

So, the overall packet delay for the proposed method $E[D_{NRT}]_{prop}$ can be expressed as

$$E[D_{NRT}]_{prop} = (1 - \gamma_R - \gamma_F)E[D_{N\_DRX}] + \gamma_R E[D_{DRX}]_R + \gamma_F E[D_{DRX}]_F$$

$$= \frac{1}{2(1-\rho)}\left[\lambda E[S^2] + (\gamma_R + \gamma_F)\left\{T_s + \frac{(n+1)T_{st}}{2}\right\}\right]. \tag{4.50}$$

Similarly, $E[W_D]$ can be estimated for the existing method as $E[W_D]_{exist\_S}$ and $E[W_D]_{exist\_L}$ for short and long DRX cycles, respectively and they can be expressed as

$$E[W_D]_{exist\_S} = \frac{T_s}{2(1-\rho)} \tag{4.51}$$

and

$$E[W_D]_{exist\_L} = \frac{T_L}{2(1-\rho)}. \tag{4.52}$$

Denoting the probability of packet arrival during short cycle and long cycle of the existing method as γ_S and γ_L, respectively, they can be expressed as

$$\gamma_S = \frac{E[T_S]}{E[T_{Total}]_{exist}} \tag{4.53}$$

and

$$\gamma_L = \frac{E[T_L]}{E[T_{Total}]_{exist}}. \tag{4.54}$$

The overall packet delay for the existing method $E[D_{NRT}]_{exist}$ can be expressed as

$$E[D_{NRT}]_{exist} = (1 - \gamma_S - \gamma_L)\, E[W_Q] + \gamma_S \left[E[W_Q] + \frac{T_S}{2(1-\rho)}\right] + \gamma_L \left[E[W_Q] + \frac{T_L}{2(1-\rho)}\right]$$

$$= \frac{1}{2(1-\rho)}\{\lambda E[S^2] + \gamma_S T_S + \gamma_L T_L\}. \tag{4.55}$$

The mean delay in service resumption after a long inactivity, without terminating the session, can be estimated as $E[W_D]$. So, it will be $E[W_D]_{prop}$ and $E[W_D]_{exist\_L}$ for the proposed and existing methods, given by (4.46) and (4.52), respectively. When $\mu \gg \lambda$, $E[W_D]_{prop} \approx E[R]_{prop}$ and $E[W_D]_{exist\_L} \approx T_L$ and then $E[W_D]$ will not vary significantly with the packet arrival rate.

4.7   Simulation

The performance has been evaluated for the proposed and existing methods numerically using simulation. The simulation implements the analytical models shown in Section 4.6 using MATLAB.

[42] gives the typical values of PDB as $PDB_{RT} = 100\ ms$ and $PDB_{NRT} = 300\ ms$. As shown in Section 4.4, the overall delay due to RLC retransmissions can be estimated as the summation of the reordering timer value and RTT. The RTT is commonly estimated as 8 ms [1]. As was shown in Section 3.4.3, the reordering timer value can take on values between 0 ms and 100 ms with 5 ms gaps or values between 100 ms and 200 ms with 10 ms gaps. Assuming all reordering timer values in the specification equally likely to occur, the mean reordering timer value can be set to 85 ms. So, the mean overall delay due to RLC retransmissions $E[D_{RTX}]$ is set to 93 ms. Using (4.3), $\bar{\zeta}$ can be estimated as 107 ms. So, the constraint in (4.5) is $\Delta_{NRT} \leq 107\ ms$. The performance evaluation was performed for three different cases for which the assumptions are shown in Table 4.1. $\mu \gg \lambda$ is used, which is usually the case. The values of T_I, T_S and T_L in Table 4.1 comply with the permissible values

in [18]. The values of P_{Data}, P_{wait} and P_{OFF} in Table 4.1 follow the UE power consumption model used in [47]-[49].

Table 4.1 Simulation assumptions for proposed and existing DRX methods

Parameter	Case 1	Case 2	Case 3
T_S	20 ms	10 ms	10 ms
T_L	320 ms	640 ms	40 ms
T_{st}	20 ms	30 ms	10 ms
n	15	21	3
T_I	10 ms		
N	1, 2, 3, 4, 5, 8, 12 and 16		
P_{Data}	500 mW		
P_{wait}	255 mW		
P_{OFF}	11 mW		
Packet arrival rate (λ)	0.05 to 0.5 packets/ms		
Service rate (μ)	100 packets/ms		

Fig. 4.3 shows that the percentage power consumption, in case 1, has significantly improved in the proposed method compared to the existing method for almost all values of packet arrival rate λ. Fig. 4.4 shows the relative DRX period $\frac{E[T_{OFF}]}{E[T_{Total}]}$ indicating that the proposed method exhibits much longer DRX period in most cases. However, the existing method with N = 1 has better power saving for a significant range of λ. The power saving is also little better for N = 2 but for a very limited range of λ. Evidently, as the packet arrival rate increases, the duration for data transfer and inactivity timer increases while terminating the DRX operation quicker and consequently, the power saving declines in all cases. [32] shows that short DRX cycles are very effective in reducing latency. This is also demonstrated by Fig. 4.5 in which the delay takes on the best values in the existing method with very high number of short DRX cycles. This delay performance is used as the benchmark to evaluate Δ_{NRT} for other cases. Compared to this benchmark, the proposed method has roughly 60 ms longer delay for λ = 0.05 packets/ms and so, Δ_{NRT} is set to 60 ms. This increase in delay can be tolerable because $\Delta_{NRT} \leq 107\ ms$. Conversely, the existing method with N = 1, has almost 120 ms longer delay for λ = 0.05 packets/ms and this increased delay may not be acceptable because $\Delta_{NRT} > 107\ ms$. Thus, repudiating the existing method with N = 1, the

proposed method remains the best in performance. The delay is found to drop almost linearly with the packet arrival rate in the proposed case. This is partly because the DRX cycle length linearly decreases in reverse time scale and a higher packet arrival rate increases the likelihood of terminating DRX operation in an earlier DRX cycle.

Fig. 4.6 shows that the percentage power consumption, in case 2, becomes much better in the proposed method for a significant range of λ, except when the existing method has N = 1, 2 and 3. The power saving is little better in the existing methods with N = 4 and 5, but for a very limited range of λ. As shown in Fig. 4.7, compared to the benchmark, the proposed method has an increase in delay $\Delta_{NRT} = 100\ ms$ for $\lambda = 0.05$ packets/ms. This increased delay can be tolerable because $\Delta_{NRT} \leq 107\ ms$. Conversely, the delay in the existing method, even with N = 4, is almost 200 ms longer than the benchmark for $\lambda = 0.05$ packets/ms. Thus, the existing method is not acceptable for small values of N as $\Delta_{NRT} > 107\ ms$. Consequently, the proposed method can be regarded as the best option.
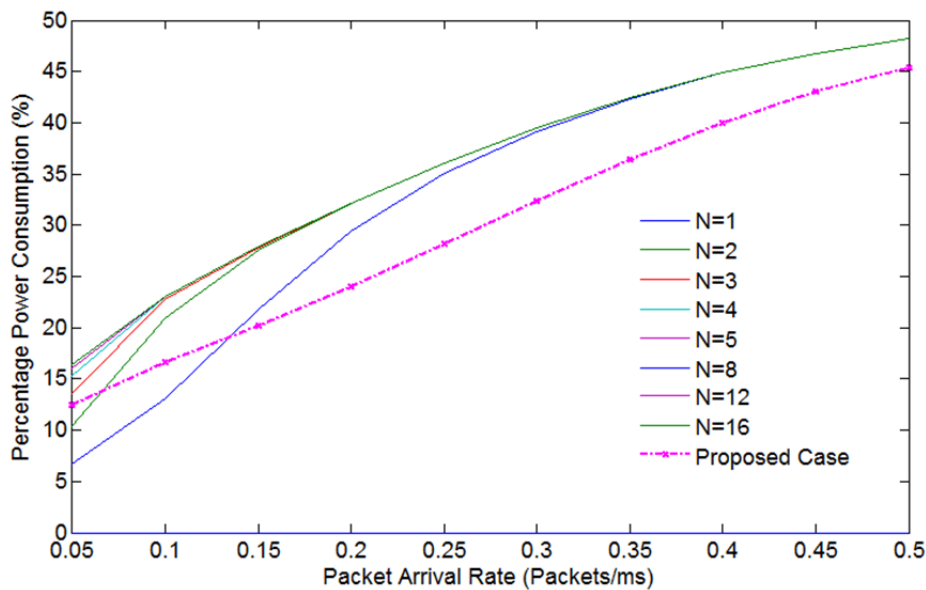
Figure 4.3: Percentage power consumption vs. packet arrival rate (case 1)

In case 3, a small value of T_L is used. In this case, as Fig. 4.8 shows, the power saving becomes better in the proposed method compared to the existing method with any value of N. Fig. 4.9 shows that all the cases here satisfy $\Delta_{NRT} \leq 107\ ms$.

The mean delay in service resumption after a long inactivity for all three cases are shown in Table 4.2 for the proposed and existing methods with $\lambda = 0.05$ packtes/ms. With the assumption $\mu \gg \lambda$, this delay will not vary significantly with the packet arrival rate and so, it is shown for a single value of λ. This delay is evidently much better in the proposed method.
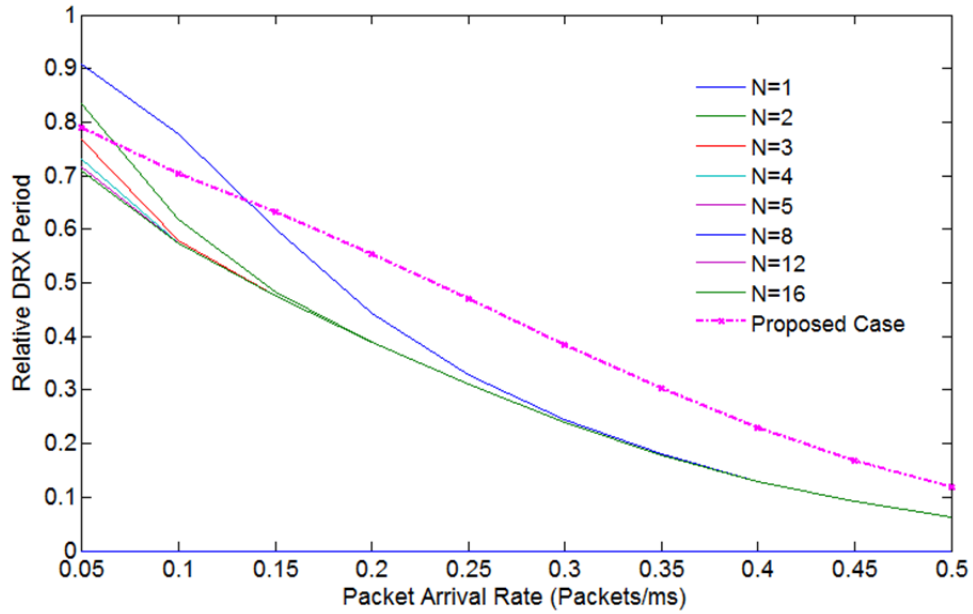
Figure 4.4: Relative DRX period $\left(\frac{E[T_{OFF}]}{E[T_{Total}]}\right)$ vs. packet arrival rate (case 1)
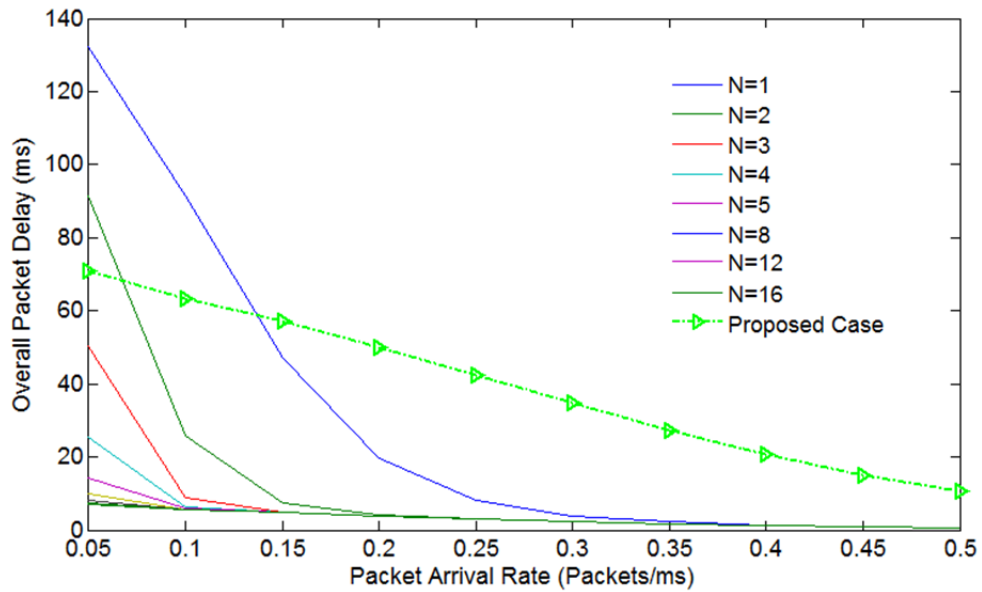
Figure 4.5: Overall packet delay vs. packet arrival rate (case 1)

Table 4.2 The mean delay in service resumption after a long inactivity

Method	Case 1	Case 2	Case 3
Existing Method	160.08 ms	320.16 ms	20.01 ms
Proposed Method	90.05 ms	170.09 ms	15.01 ms
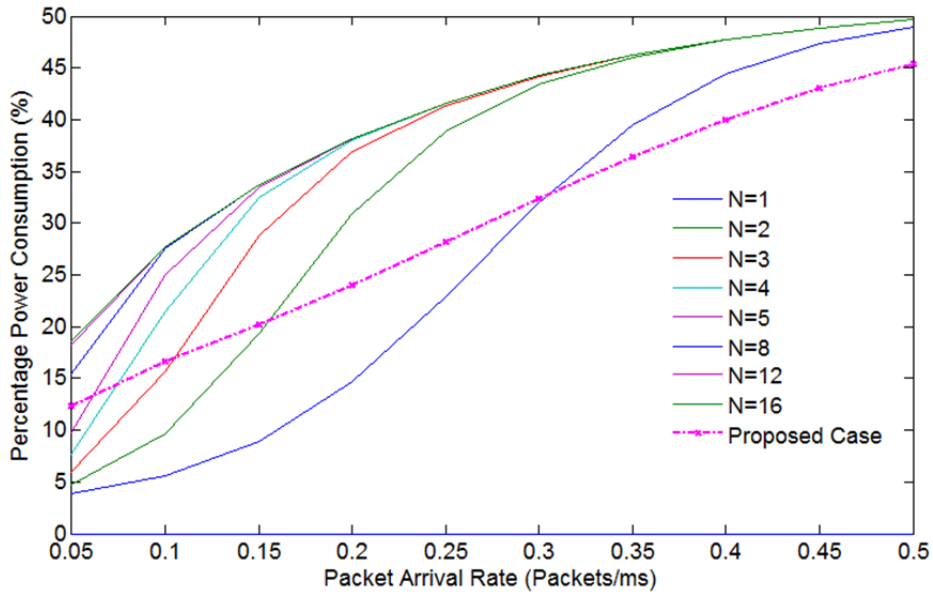
Figure 4.6: Percentage power consumption vs. packet arrival rate (case 2)
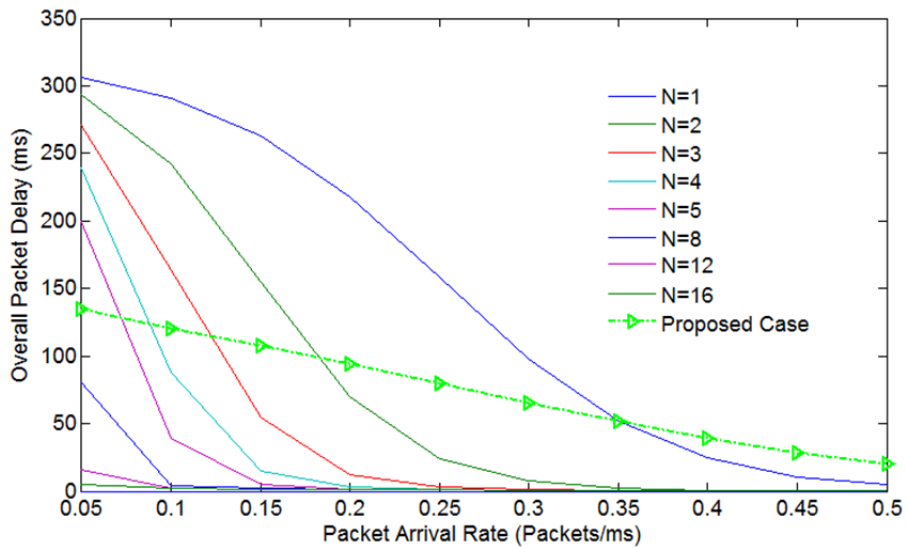
Figure 4.7: Overall packet delay vs. packet arrival rate (case 2)

The above numerical examples demonstrate that a significantly better power saving is achieved in the proposed method with permissible increase in packet delay. The existing method may often attain an even better power saving setting very small values of N but its associated delay might exceed the tolerable limit then. The delay in service resumption after a long inactivity is always better in the proposed method.
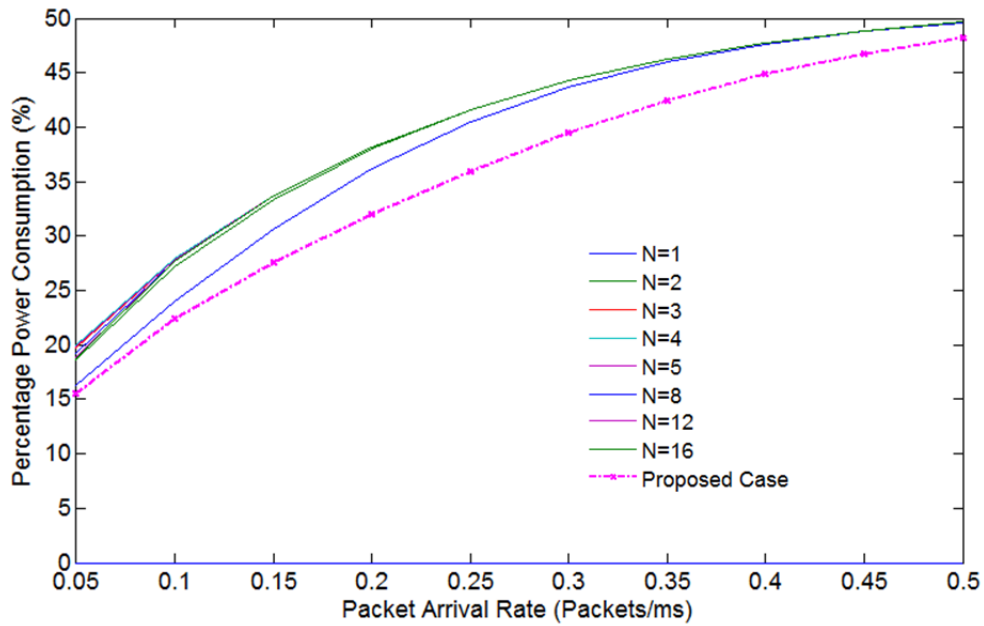
Figure 4.8: Percentage power consumption vs. packet arrival rate (case 3)
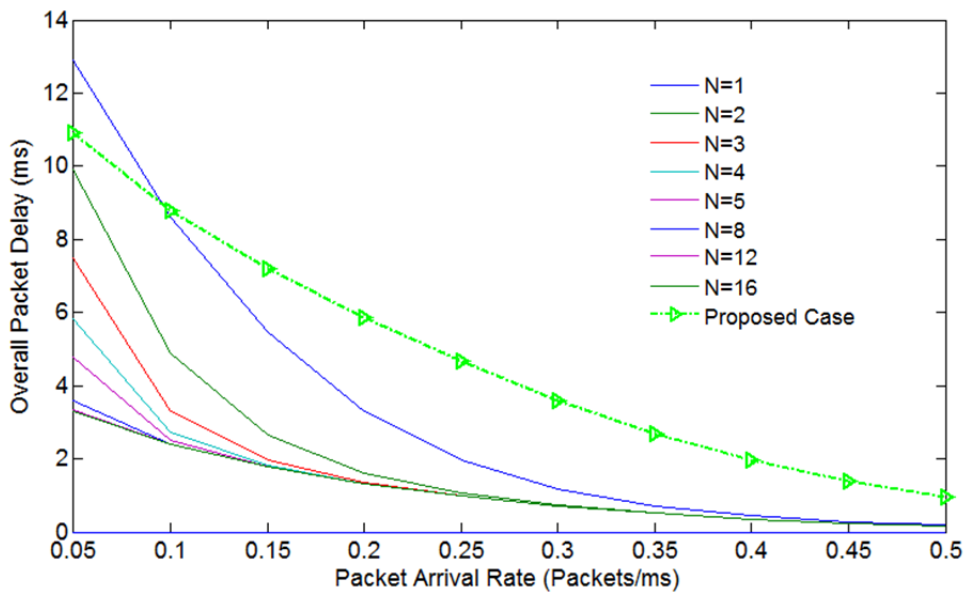
Figure 4.9: Overall packet delay vs. packet arrival rate (case 3)

4.8   Analysis of Results

The major factor in the difference of packet delay between RT and NRT traffic is shown as the RLC retransmissions. The RLC retransmission takes on a delay, which is the summation of the reordering timer value and the RTT. Taking the mean value of the possible values of reordering timer, it is set to 85 ms in the simulation. This allows RLC retransmission to take on a delay of 93 ms. However, the permissible difference in delay

between RT and NRT traffic is much higher than 93 ms. This is because the users do not care much about the delay between two packets for NRT traffic. According to the specifications, the permissible difference in delay between RT and NRT traffic is estimated as 200 ms. Thus, the additional 107 ms becomes allowable increase in delay for NRT traffic. This exemplifies that the a significant delay becomes available in hand as an allowable increase in delay for NRT traffic and this delay is exploited to improve the DRX power saving in the proposed scheme.

The simulation results for the existing method can be considered similar with RT traffic when the value of N is around its maximum possible value 16. This is because RT traffic can use such high value of N. On the other hand, NRT traffic can use low value of N but not too low to make the delay intolerable. The simulation results show that the lowest possible values of N for the existing method with sufficiently small delays offers much higher power consumption compared to the proposed method. The proposed method gradually increases DRX cycle length to save power. The proposed method also gradually decreases DRX cycle length in order to reduce the mean delay in service resumption after a long inactivity. The simulation results demonstrate a significant reduction in delay in this service resumption.

4.9 Discussion and Conclusion

The tradeoff between power saving and delay for DRX operation applies to all types of traffic. Since NRT applications are delay tolerant, to improve power saving, a DRX method is proposed for use when only NRT applications are running with no RT applications active. The proposed method avoids the estimation of current traffic.

Typically, NRT applications use RLC AM whereas RT applications use RLC UM. The RLC AM provides reliable in-sequence delivery of data using ARQ. Since multiple HARQ processes run in the MAC layer, there will be additional delays for transport blocks, which are retransmitted in the HARQ protocol. Also, there can be difference in the propagation delay for PDUs. Thus, the RLC can receive out of sequence PDUs. If the receiving RLC entity finds a gap in the sequence of the PDUs, it starts a reordering timer. When the timer expires without the arrival of the missing PDU, the receiving entity infers that HARQ has failed to recover the PDU or it has been lost. At this juncture, RLC UM does not attempt to recover the lost PDUs but RLC AM does. The receiving RLC AM entity sends a status report that includes positive acknowledgement for correctly received PDUs and

negative acknowledgement for fully and partially missing PDUs. Then the transmitting RLC AM entity retransmits the missing data. The overall delay in the correction of data can be estimated as the summation of the reordering timer value and round trip time (RTT). The RTT includes round trip propagation delay and processing delay at both ends. The reordering timer value is configured by the eNodeB. Thus, the RLC AM incur some delay and this occurs for NRT traffic.

The buffer size is relatively smaller for RT data and packets are lost when the queue exceeds this size. The NRT buffer size is larger to avoid packet losses and some queue analyses even consider it infinite. RT traffic has a stringent delay requirement, while some level of packet loss rate can be acceptable. Conversely, NRT traffic is error-sensitive but the transmission delay can be larger. Thus, the NRT traffic permits a limited amount of delay essentially with no adverse impact on the user's experience. A part of the additional delay, permissible for NRT traffic in comparison with RT traffic, is consumed due to RLC retransmissions. After this consumption, it can be shown that some delay is left over, which can be allowed to NRT traffic. The proposed method increases delay deliberately within permissible limits leading to a significant improvement of power saving. The proposal is applicable only when NRT applications are running with no RT applications active. In the proposed method, the MAC layer keeps varying the DRX cycle length in a simple fashion. The RRC layer at eNodeB configures parameters related to the variation in DRX cycle length considering the operator's choice and an overall estimate of the potential traffic. Analytical models have been derived to evaluate both the proposed and the existing methods. The numerical examples exhibit success in power saving in the proposed method. A new factor has been defined for a more practical estimation of power saving. The proposed method also attains less mean delay in service resumption provided that there has been a long inactivity.

The user, in the reality, can be mobile and he can even move at a high speed. As he moves, he passes different cells and so, the serving cell must be updated to support the mobility. This can pose some difficulties, which need to be addressed. There are even more challenges when there are cells of different capacities. The next chapter discusses these challenges as well as how to address them with a goal for better performances.

Chapter 5

Enhanced Control over Cell Switching for Speedy Users in HetNets

The HetNets have promising features but it involves new challenges to support seamless mobility, especially for high speed users. The improvement of the performance lies in the achievement of better control at RRC layer over the cell switching and proper setting based on the particular scenario. This chapter delves into the requirement of flexible and adaptive control over cell switching procedure for speedy users and a method is proposed to achieve better controllability leading to performance improvement.

5.1   Introduction

The general procedure for mobility support in LTE was explained in Section 2.8. In HetNets, due to the increased number of cells, a high number of cell switching needs to be performed to support seamless mobility, especially when the user has a high speed as shown in Fig 5.1. But it is difficult to support the seamless mobility of UEs in a HetNet scenario and handovers may fail [51]. Since the coverage ranges of eNodeBs are different in HetNets, the handover performance of a UE may get significantly degraded, especially in scenarios where high speed UEs traverse through small cells. The result may be handover failures (HFs), radio link failures (RLFs), as well as ping-pong events [52]. The HF, RLF and ping-pong events were explained in Section 2.8. [53], [54] demonstrate that HFs and ping-pong events can be serious problems in the case of HetNets because of small cells. The improvement of the performance lies in the achievement of better control at RRC layer over the cell switching and proper setting based on the particular scenario.

5.2   Identification of Difficulties in Mobility Support in HetNets

Currently, the handover procedure and its incorporation of parameters are the same for a traditional macro-cellular network and for HetNets. In scenarios with only macrocells, HFs and ping-pong events can be typically avoided due to large cell sizes [55]. However, the

deployment of a large number of small cells in HetNets increases the complexity. Using the same set of handover procedure and parameters of a traditional macro cellular network will degrade the mobility performance of the UEs for a HetNet scenario [53]. The proper cell switching is a challenge in HetNets for speedy users because of various reasons as shown below.

1. Section 2.7.2 explains that overlaying of macro cell and small cells in a HetNet results in severe interference, in particular for cell edge users of small cells. Thus, the SINR can be very poor at cell edge of small cells even for users who are not moving.

   The handover dragging effect is particularly severe for small cells or in a HetNet scenario with high speed UEs [56]. This handover dragging effect worsens SINR at cell edge. However, as shown above, SINR is already poor at cell edge even for immobile users in HetNets. Thus, for speedy UEs, the SINR can fall too low degrading cell edge performance severely. In the worst case, there can be RLF and HF.
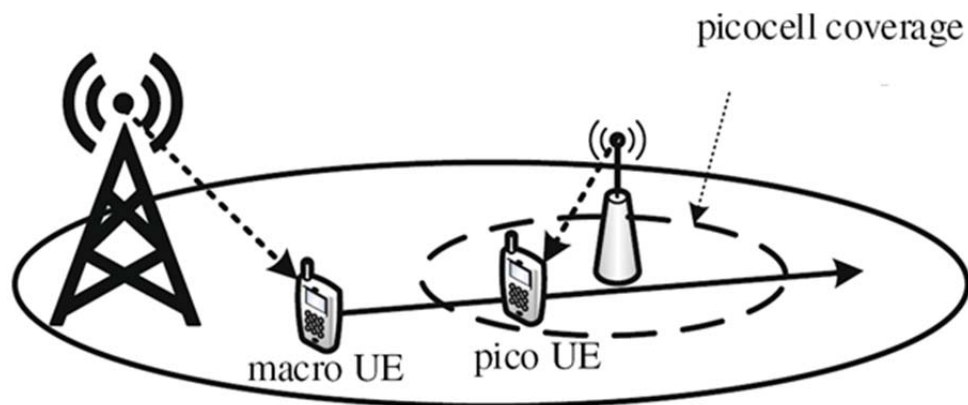
Figure 5.1: A speedy UE is handed over from macro eNodeB to pico eNodeB

2. As shown in Section 2.8.2, the signal measurements obtained at a UE from the neighboring eNodeBs are reported by the UE to its serving eNodeB and the handover decision is made by the serving eNodeB. In HetNets, such measurement reporting by the UE may not be as quick as it is required considering the small cell size and high speed of the users. This may result in handover failure (HF) problems for the speedy users [57]. This is illustrated in Figure 5.2. Similar problem can also occur for cell reselection.

3. In current 3GPP standards, the RRC layer of the UE applies scaling of cell switching parameters in high and medium mobility state of the user as shown in Section 2.8.3.

Thus, scaling is performed at only two discrete steps for a speedy user. With only these two steps, the existing method tends to take longer delay to execute cell switching. This causes the UE to move far inside the new cell before the new cell takes over. The minimum received power P_{R_Min}, defined in Section 2.8.2, falls very low. The very poor radio link quality from the old cell results in a highly degraded data rate and in the worst situation, there can be RLF and HF.
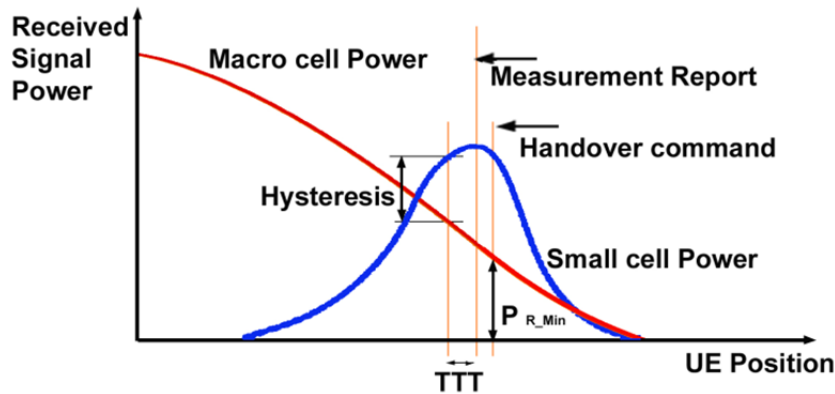
Figure 5.2: Illustration of the problem in handover for speedy users in HetNets

4. The receiver sensitivity for the eNodeB and the UE is required to support P_{R_Min} values in uplink and downlink, respectively. In the existing method, with only two steps in scalling, P_{R_Min} takes on widely varying values at different velocities. This may pose difficulty to match between the receiver sensitivity and P_{R_Min} in the system. It may be noted that P_{R_Min} falls extremely low at very high user velocities and in this case, the very low received power persists for a very short period before cell switching. So, in the existing method, it may not be worthwhile to support the lowest P_{R_Min} but without the support, the service can be degraded or disrupted.

5. In a HetNet scenario, the existence of high number of cells can lead to a lot of overlapping in cell coverages. Thus, the cell boundaries and target cells for handover can be far from clear. Also, there can be multiple cell borders and different target cells at different times, due to fast variation of SINR from each cell [56]. These can cause difficulties in successful cell switching.

6. The existing method relies merely on the history of cell switching for the decision of scaling. But this history is not always a good indication of the user speed. For example, a user at low velocity may take a tangled pathway within overlapped cells and undergo frequent cell switching. Then the existing method applies high scaling,

which is not justified because a sufficient margin is required to avoid any ping-pong effects. Here, if the high scaling was not applied, the UE would not go far inside another cell. Thus, the possibility of ping-pong effects is unnecessarily made higher.

7. The HetNet along with its small cells are mostly expected in urban areas. It is not always true that high speed UEs do not exist in urban areas with small cells, due to frequent red light stops. Because, in modern cities, highways can go through cities and cars can drive at high speeds without making stops. Also, bullet trains can pass through cities at high speeds, as shown in Figure 5.3 [56].

Figure 5.3: High speed users in urban areas

8. In HetNets, the cell switching parameters need to be adjusted independently per neighbor cell considering the different characteristics of the neighbor cells. This is because the handovers between macro-cells and small cells have different characteristics compared to handovers between macro-cells. For handovers from small cells to macro-cells, the received power from the small cell drops dramatically compared to macro-cells. So, the UE suffers inter-cell interference from the macro-cell earlier. If the handover parameters are stringent, too late handovers can occur. For

87

handovers from macro-cells to small cells, the received power from the small cell increases dramatically compared to macro-cells, the UE suffers inter-cell interference from the small cell earlier. Like the small cell to macro-cell scenario, too late handovers can occur if the handover parameters are stringent.

9. The cell switching conditions should be highly flexible and adaptive to fit into the current loading conditions. The cell range expansion (CRE), which was explained in Section 2.7.1, can vary its extent depending on loading conditions. Also, the cell switching depends on the traffic usage pattern as explained below.

Section 3.6 shows that the UE decodes PDCCH for its resource allocation information. For data channels, HARQ can help data when the BLER is around 10%. But HARQ cannot be applied to PDCCH. Thus, if the BLER on PDCCH exceeds 10%, the UE fails to receive response from the eNodeB. This causes HF.

The load on PDCCH depends on the type of application. A small number of high data-rate users, using applications like FTP transfer gets PDCCH lightly loaded. On the other hand, the PDCCH gets heavily loaded if there are a large number of low-rate users in the cell, using applications like VoIP. When PDCCH is lightly loaded, its different parts can distributed for use among the neighboring cells. This high reuse ratio of PDCCH produces little interference, allowing better SINR. But when PDCCH is heavily loaded, the interference scenario will be more like reuse ratio 1, then PDCCH suffers from low SINR. Thus, the cell switching parameters should adapt to the traffic usage pattern and this adaptability is not enough strong in the current standards [56].

10. In HetNets, the downlink coverage of a pico base station is much smaller than that of a macro base station. This is not the case for uplink where the strength of the signal received from a UE and all UEs have almost equal transmit power capabilities. Hence, the uplink coverage of all the base stations is similar and the uplink handover boundaries can be determined based on channel gains. This can create a mismatch between downlink and uplink handover boundaries, and make the selection of an appropriate cell for service more difficult in heterogeneous networks, compared to homogenous networks, where downlink and uplink handover boundaries are more closely matched [58].

5.3   Related Work

Handover parameters (e.g. time-to-trigger (TTT), hysteresis threshold, etc.) are optimized in [59] to achieve robust and seamless mobility of UEs in a HetNet scenario. The conclusions are aligned with the HetNet mobility performance evaluations in [53], showing that HetNet mobility performance strongly depends on the cell size and the UE speed. In [60], mobility performance of UEs is evaluated in the co-channel small cell networks scenario; when the density of the small cell increases, switching off the macro cell is shown to provide seamless mobility for the low speed UEs, while it degrades the handover performance for the high speed UEs [61]. Mobility state estimation is performed in [62] to estimate the velocity of the UEs and thereby keeping the high speed UEs to macrocells and low speed UEs are offloaded to picocells, thereby enhancing the handover performance of the UEs. [52] proposes a context-aware mobility management (MM) procedure for HetNets, in which the base stations jointly learn their long-term traffic loads and optimal cell range expansion and schedule their UEs based on their velocities and historical data rates. [63] shows that the adjustment with user velocity with higher level of granularity improves the performance. There are many proposals related to adjustment of cell switching parameters so far. But there is still deficiency in the controllability that is required for speedy users in HetNets.

5.4   Proposed Scheme

The problems in proper cell switching in HetNets for speedy users were explained in Section 5.2. In order to solve such problems, mobility parameters in each cell need to be dynamically optimized according to cell traffic loads, coverage areas of different cells, and velocities of the UE [52]. Section 2.8.3 explains the scaling of cell switching parameters based on user speed in the current standards. But the existing standards or current proposals, shown in Section 5.3, do not provide sufficient controllability on cell switching parameters in HetNets for speedy users. The cell switching conditions should be more flexible and adaptive so it can properly fit into the current status considering coverage, loading and user speed. A more flexible and more adaptive way to perform scaling is proposed such that the cell reselection or handover occurs at a more appropriate moment.

It is assumed that the user moves along trajectory 1 in Fig. 5.4 at velocity v. Here, trajectory 1 is a straight line connecting the macro and pico eNodeBs. D_S and D_T are the

distance of the cell border between the source cell and the target cell. It is possible to estimate v from Doppler spread reliably for different delay spreads from downlink OFDM signals. It is well known that the user velocity v and the maximum Doppler shift f_d are related for a fading signal as

$$v = f_d \lambda = \frac{c f_d}{f_c}. \tag{5.1}$$

Here, f_c is the signal frequency, λ is the wavelength and c is the speed of light. A reliable estimate of f_d is required to determine v. The velocity is estimated in [64] using the relationship given by

$$v = \frac{c f_d}{f_c} = \frac{2.405 c}{2 \pi f_c T_S \hat{l}_0}. \tag{5.2}$$

Here, T_S is OFDM symbol duration and $\hat{l}_0$ is the zero crossing point of the estimated covariance function, $\varphi_t(l)$ of the received signal at a certain carrier in the frequency domain, where l represents the difference in time. The velocity estimation is shown almost independent on the SINR in [64] and so, it can be used even when the radio link is poor.

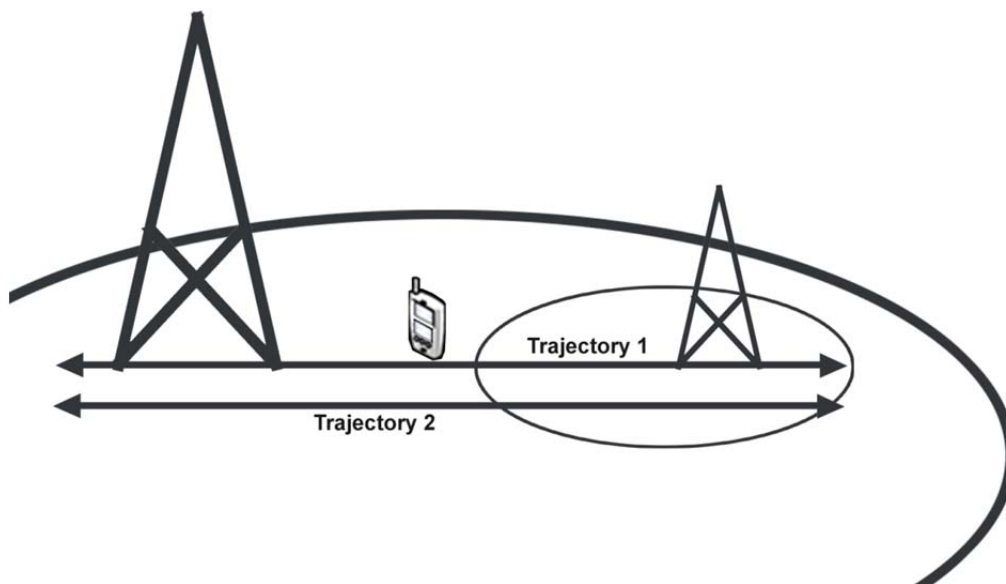**Figure 5.4:** The trajectory along which the user moves

In the proposal, the UE estimates its current velocity based on the Doppler spread using the cell-specific reference signals, transmitted by the eNodeB. The UE can apply one of the methods for estimation of Doppler spread for OFDM signals, suggested in [64] [65][66]. Then the UE uses a combination of the number of cell reselections or handovers in the recent

history and the Doppler spread based velocity estimation for computation of the scaling factor. An eNodeB controlled scaling of the hysteresis margin is proposed to establish a set point. Then the UE performs adjustment of timer period as an overlay around the set point at the discretion of the UE itself. The proposal uses normal mobility state and high mobility state only leaving the medium mobility state out. A smooth scaling is performed when the UE is in high mobility state and the normal mobility uses no scaling. The proposal is explained below for cell reselection and handover.

5.4.1   Cell Reselection

Section 2.8.1 shows that the ranking method of serving and neighbor cells to evaluate the cell reselection condition is that

the rank of the serving cell, $R_S = Q_{meas,S} + Q_{hyst}$

and the rank of the target cell, $R_T = Q_{meas,T} - Q_{offset}.$ (5.3)

The condition for the triggering of the $T_{reselectionRAT}$ timer is assumed to be met at a distance d from the cell border between the source cell and the target cell. According to the log distance path loss model, at d, the received power from the serving cell and the target cell can be expressed in dB as $\overline{Pr_S}(d_0) - 10n \, log\left(\frac{D_S+d}{d_0}\right)$ and $\overline{Pr_T}(d_0) - 10n \, log\left(\frac{D_T-d}{d_0}\right)$ respectively, where n is the path loss exponent for the multipath environment. Thus, ignoring quantization in power level calculation, (5.3) can be shown as

$$10n \, log\left(\frac{D_S + d}{d_0}\right) - 10n \, log\left(\frac{D_T - d}{d_0}\right) = Q_{hyst} + \overline{Pr_S}(d_0) - \overline{Pr_T}(d_0)$$ (5.4)

where Q_{offset} has been set to zero assuming zero bias to the target cell. So,

$$4.343n\left[ln\left(1 + \frac{d}{D_S}\right) - ln\left(1 - \frac{d}{D_T}\right) + ln\left(\frac{D_S}{D_T}\right)\right] = Q_{hyst} + \Delta_{Pr}(d_0)$$

where $\Delta_{Pr}(d_0) = \overline{Pr_S}(d_0) - \overline{Pr_T}(d_0)$. $\Delta_{Pr}(d_0)$ will be positive when the UE is moving from macro cell to pico cell and it will be negative when the UE is moving from pico cell to macro cell. Using Taylor series expansion and ignoring higher order terms assuming $d \ll D_S$ and $d \ll D_T$,

$$4.343n\left(\frac{1}{D_S} + \frac{1}{D_T}\right)d + 4.343n \, ln\left(\frac{D_S}{D_T}\right) - \Delta_{Pr}(d_0) = Q_{hyst}.$$ (5.5)

$4.343 n \ln\left(\frac{D_S}{D_T}\right) - \Delta_{Pr}(d_0)$ is a constant for a particular scenario. To attain smooth scaling, d is shifted linearly towards the serving cell as the user velocity v increases and for this purpose, Q_{hyst} can be scaled as

$$4.343 n \left(\frac{1}{D_S} + \frac{1}{D_T}\right) d + 4.343 n \ln\left(\frac{D_S}{D_T}\right) - \Delta_{Pr}(d_0) = Q_{hyst} - k'v. \tag{5.6}$$

Here, k' in (5.6) is proposed to be eNodeB controlled using the parameter $Q_{hyst}SF$ where $Q_{hyst}SF$ is negative in dB value and so,

$$4.343 n \left(\frac{1}{D_S} + \frac{1}{D_T}\right) d + 4.343 n \ln\left(\frac{D_S}{D_T}\right) - \Delta_{Pr}(d_0) = Q_{hyst} + Q_{hyst}SF.kv .$$

The UE determines kv as a factor proportional to v and the factor will be called *Vel_factor_CR* as can be expressed as follows.

$$4.343 n \left(\frac{1}{D_S} + \frac{1}{D_T}\right) d + 4.343 n \ln\left(\frac{D_S}{D_T}\right) - \Delta_{Pr}(d_0)$$
$$= Q_{hyst} + Q_{hyst}SF.Vel\_factor\_CR \tag{5.7}$$

The UE can, of course, find itself in a homogenous network and then it moves within only macro cells. In this case, assuming all macro cells to have radius D and equal transmit power at eNodeBs, (5.8) will take the simple form,

$$\frac{8.686 n}{D} d = Q_{hyst} + Q_{hyst}SF.Vel\_factor\_CR . \tag{5.8}$$

Then a combination of the number of cell reselections in the last T_{CRmax} period, N_{cr} and Doppler spread based velocity estimation is used for the determination of *Vel_factor_CR* as follows.

$$Vel\_factor\_CR = (1 - \alpha)N_{cr\_filtered} + \alpha\beta f_d \tag{5.9}$$

$$N_{cr\_filtered} = (1 - \gamma)N_{cr} + \gamma N_{cr\_last} \tag{5.10}$$

$$\gamma = 1/2^\eta \tag{5.11}$$

Here, $N_{cr\_filtered}$ represents the filtered value of N_{cr} and it is updated in every T_{CRmax} period. N_{cr_last} gives the last filtered value of N_{cr}. η is the filter coefficient, which can be varied between 0 and 6. It is expected that the impact of N_{cr} is more important than N_{cr_last} and η will generally be high. (5.11) allows better granularities for the impact of N_{cr} for higher

values of η. βf_d represents the velocity estimated from Doppler spread along with a gain factor. The value of β should be chosen to yield the same order of values for $N_{cr\_filtered}$ and βf_d. It may be noted that β decreases with f_c and so, it will vary with the operating band. α is used to control the relative influence of the number of cell reselections and Doppler spread.

Finally, $Q_{hyst\_}scaled$ is computed as follows, which the UE applies as the hysteresis margin.

$$Q_{hyst\_}new = Q_{hyst} + (\ Q_{hyst}SF \times Vel\_factor\_CR) \tag{5.12}$$

$$Q_{hyst\_}scaled = [Q_{hyst\_}new]_{round\_n} \tag{5.13}$$

$[.]_{round\_n}$ represents rounding to the nearest value of n where n is an integer. This allows the implementation of the scaled hysteresis margin to comply with the existing discrete steps so that it does not increase complicacy substantially.

Once the scaling of the hysteresis margin with the user speed establishes a set point, the UE additionally scales $T_{reselectionRAT}$ for the purpose of adjustment around the set point. Once the timer is triggered at a distance d, the timer $T_{reselectionRAT}$ will run over an additional distance d' before triggering cell reselection where,

$$d' = v.T_{reselectionRAT} \tag{5.14}$$

In order to keep d' reasonably constant, $T_{reselectionRAT}$ is linearly varied with v as

$$T_{reselectionRAT\_new} = T_{reselectionRAT} \cdot \frac{SSSF}{Vel\_factor\_CR}\,. \tag{5.15}$$

Here, SSSF is used as the proportionality constant and the UE itself determines SSSF from a range of values provided by the eNodeB.

$$T_{reselectionRAT\_scaled} = \left[T_{reselectionRAT\_new}\right]_{round\_0.125n} \tag{5.16}$$

The UE applies $T_{reselectionRAT\_scaled}$ as the scaled timer period and it is rounded to its nearest multiple of 125 msec in order to comply with the existing implementation. In addition to $Q_{hyst}SF$, the eNodeB needs to send α, β, η and a range of values for SSSF to the UE via system information message. The proposal obviates the need for conveying N_{CR_M} since the medium mobility state is removed.

5.4.2   Handover

The proposed method of scaling for handover is similar to the proposition for cell reselection explained is Section 5.4.1. A factor proportional to the user velocity Vel_factor_HO is first calculated using the number of handovers in last T_{CRmax} period, N_{ho} and the Doppler spread as follows.

$$Vel\_factor\_HO = (1 - \alpha)N_{ho\_filtered} + \alpha\beta f_d \qquad (5.17)$$

$$N_{ho\_filtered} = (1 - \gamma)N_{ho} + \gamma\, N_{ho\_last} \qquad (5.18)$$

$$\gamma = 1/2^{\eta} \qquad (5.19)$$

The existing method for handover does not scale hysteresis with the user speed as shown in Section 2.8.3. Thus, the distance d_{HO} at which the conditions triggering the timer are met, remains independent of the user velocity. The scaling of hysteresis is performed in order to shift d_{HO} linearly towards the serving cell as v increases and thus, a set point for scaling is established. Thus,

$$Hysteresis\_new = Hysteresis + (Hyst - SF \times Vel\_factor\_HO\,). \qquad (5.20)$$

Here, $Hyst - SF\ [dB]$ is controlled by the eNodeB and it is negative in dB value. The UE applies $Hysteresis\_scaled$ as the scaled hysteresis and it is rounded to its nearest multiple of 500 msec as

$$Hysteresis\_scaled\ = [Hysteresis\_new]_{round\_0.5n}. \qquad (5.21)$$

The distance d''over which the entering conditions for a particular event need to be met can be expressed as,

$$d'' = v.TimeToTrigger. \qquad (5.22)$$

In order to keep d'' reasonably constant, $TimeToTrigger$ is linearly varied with v as

$$TimeToTrigger_{new} = TimeToTrigger.\frac{SSSF}{Vel\_factor\_HO}\,. \qquad (5.23)$$

The UE determines SSSF from a range of values provided by the eNodeB. The UE applies $TimeToTrigger_{scaled}$ as the scaled timer period and it is rounded to its nearest multiple of 128 msec in order to comply with the existing implementation as

$$TimeToTrigger_{scaled} = [TimeToTrigger_{new}]_{round\_0.128n} \qquad (5.24)$$

94

The UE applies $TimeToTrigger_{scaled}$ as the scaled timer period and it is rounded to its nearest multiple of 128 msec in order to comply with the existing implementation.

The eNodeB needs to send α, β, η, Hyst-SF and a range of values for SSSF to the UE via RRCConnectionReconfiguration message.

5.4.3   *Controlling the Scaling*

The scenarios in HetNets may vary widely as a result of differences in transmit power, coverages of macro and pico cells, traffic loading in macro and pico cells, position and traffic pattern of highways and other streets with respect to the coverage of macro and pico cells, profile of the user speed, direction of the user, operating frequency, multipath environment and fading condition, QoS of data services, and so forth. The proposed method provides high flexibility and adaptability in controlling the scaling for speedy users and thus, it addresses the challenges depicted in Section 5.2. It attempts for better optimization based on the current conditions.

In the proposed method, the eNodeB configures Q_{hyst}SF or Hyst-SF value in order to establish a set point for scaling for cell reselection or handover, respectively. Thereafter, the UE itself performs overlay of scaling around the set point using SSSF. For this purpose, the UE constantly monitors cell reselection or handover and adjusts SSSF values. The advantage of adjustment by the UE itself is that it needs no feedback and thus, it is quick and it requires no overhead. Besides, the UE controlled overlay can function as fine-tuning of the scaling.

Increasing the absolute values of Q_{hyst}SF or Hyst-SF reduces hysteresis margin and thus, yields less stringent cell switching conditions and a quicker cell switching. On the other hand, increasing the value of SSSF increases timer period and thus, yields more stringent cell switching conditions and a delayed cell switching. Simultaneous control from both ends, using Q_{hyst}SF or Hyst-SF by the eNodeB and using SSSF by the UE, may often reduce the controllability. Therefore, the SSSF values must be within a range of values, provided by the eNodeB.

The proposed method uses a combination of the number of cell reselections or handovers in the recent history and the Doppler spread based velocity estimation for computation of the scaling factor. The rationale behind incorporating Doppler spread is explained below.

1. One between the number of cell switching and Doppler spread can be found better indicative of the requirement of scaling than the other depending on the scenario in HetNets.

2. Since the scaling with speed reduces the cell switching delay, it increases the possibility of ping-pong effects. The inclusion of Doppler spread ensures that the velocity is really high when high scaling is applied and the chances of ping-pong effects are thus mitigated.

The optimization needs to be performed and updated adjusting the values of $Q_{hyst}SF$ or Hyst-SF, SSSF, α and η. The optimization should be performed more carefully in the case of handover due to the possibility of ongoing data transfer. The optimization may use the following guidelines.

1. α is used to control the relative influence of the number of cell reselections or handovers and Doppler spread. If the UE finds that its own velocity, estimated from Doppler spread, keeps varying widely, a high α may be used because the Doppler spread now better indicates the current state of user and thus, it better estimates the likelihood of imminent cell switching. A low α may be used if there are frequent red light stops. In this case, although the users have low velocity, they have chances of ping-pong effects because they may stop or turn around [67]. Low α may also be used when the estimation of Doppler spread in the physical layer is not very reliable.

2. $Q_{hyst}SF$ or Hyst-SF and SSSF values are adjusted based on the number of RLFs, HFs and ping-pong events. This adjustment may use the cost function suggested in [68]. [68] suggests that metrics are first calculated from the ratio of the number of too early handover, too late handover or ping-pong events and the total number of handover over certain period. Then the cost function is derived applying separate weight factors to each of the metrics. However, this method may be too slowly adaptive for speedy users in HetNets. Therefore, we suggest additional adjustment by the UE using SSSF. The UE can adjust SSSF whenever it detects too early handover, too late handover or ping-pong events. Section 2.8.2 shows how too early handover and too late handover can be detected. Ping-pong events can be detected from the identity of the recently serving cells.

3. The UE estimates proper cell switching delay considering its own velocity, estimated from Doppler spread. The UE may increase or decrease SSSF when the cell switching delay seems lower or higher than the expected values, respectively.

4. Evidently, the proposed method applies smooth scaling as opposed to the only two steps for speedy users, specified in existing standards. Thus, P_{R_Min} does not fall too low. Also, P_{R_Min} becomes more independent of the user velocity. The stability of P_{R_Min} is helpful because the receiver can be designed accordingly with proper sensitivity.

5. When the UE finds that its own velocity, estimated from Doppler spread, exceeds a threshold value v_{th}, it sets SSSF to a very high value $SSSF_{high}$. This will preclude switching to pico cells frequently and cell switching will occur only between macro cells. Thus, the number of cell switching will not go too high. v_{th} and $SSSF_{high}$ are to be configured by the eNodeB.

6. In the case of handover, the respective eNodeBs are aware whether the UE is moving from macro cell to pico cell or from pico cell to macro cell. The eNodeB can send a quick notification to decrease SSSF if the UE is moving from macro cell to pico cell and vice versa.

7. The filter coefficient η can be used to control the relative influence between the recent and older number of cell switching. The filtering removes sudden errors in the estimation. An example of such error can be that a highway may take a turn around the eNodeB and the user at a high velocity may not undergo cell changes for a while.

The proposed scheme attempts to overcome the limitations of the existing method while increasing complicacy in implementation and signaling overhead to some extent.

5.5 Simulation

A MATLAB based simulation has been performed and the simulation environment is limited to partially implement the proposed method. A strong system level simulator could properly demonstrate the achievements of the proposed method but such a simulator was not accessible. The proposed method for cell reselection and handover was simulated and compared with the existing one. The simulation assumes the user is moving from macro cell to pico cell along the two straight line trajectories shown in Fig. 5.4. Trajectory 2 is a straight line parallel with Trajectory 1 at 200m distance. The simulation assumptions shown in Table 5.1 are used. Here, the assumed parameter values comply with the specified values when certain specifications are available. It may be noted that some of the parameters in Table 5.1 were explained in Section 2.8.

For simplicity, N_{cr_last} and N_{ho_last} were assumed to be equal to the average number of cell reselections or handovers in the last T_{CRmax} in which the velocity of the user is the immediate lower one in the velocity array. f_d is calculated as $f_d = v/\lambda$ and λ is calculated as c/f based on the operating downlink frequency. A single value of α is used in all simulations. This is because a change in α would not show any difference in the results in the present simplified simulation environment where both the number of cell switching and Doppler spread always have perfect linear variation with the user speed.

Table 5.1 Simulation assumptions for speedy users in HetNets

Parameter	Value
Operating frequency (DL)	1700 MHz
Separation between trajectory 1 and 2	200 m
User velocity (constant)	10, 20, 30, …,150 km/hr
User direction	Moving from macro cell to pico cell
Path loss model	Okumura-Hata
EIRP at macro eNB	40 dBm
EIRP at pico eNB	23 dBm
D_S	500 m
D_T	300 m
Antenna height at eNB	30 m
User antenna height	1 m
T_{CRmax}	4 sec
$T_{reselectionRAT}$	6 sec
Q_{offset}	0 dB
Q_{hyst}	10 dB
SF-High (Q-HystSF)	4 dB
SF-Medium (Q-HystSF)	8 dB
SF-High (SSSF)	0.75
SF-Medium (SSSF)	0.25
$Q_{hyst}SF$ (proposed case)	-0.2, -0.4, -0.6 or -0.8 dB
Reporting method for HO	Event-triggered with event A3

Table 5.1 Simulation assumptions for speedy users in HetNets (Continued)

Parameter	Value
CellIndividualOffset for neighbor cell	0 dB
CellIndividualOffset for Serving Cell	0 dB
a3-Offset	2 dB (IE value, 4)
Hysteresis	8 dB (IE value, 16)
ReportInterval	0.24 sec
No. of reports triggering HO	4
Hyst-SF (proposed case)	-0.4 or -0.7 dB
TimeToTrigger	2.56 sec
N_{CR_M} (existing case)	3
N_{CR_H} (existing case)	7
N_{CR_H} (proposed case)	2
SSSF (proposed case)	0.5, 1, 1.5 or 2
α	0.4
β	4
η	3

5.5.1 *Results for Cell Reselection*

The lowest received power, P_{R_Min} is calculated as the signal strength from the old serving cell at the moment of cell reselection. Fig. 5.5–5.9 show P_{R_Min} vs. user speed profile for different values of $Q_{hyst}SF$ and SSSF comparing existing and proposed methods. The user moves along trajectory 2 in Fig. 5.7 while trajectory 1 is used in all other figures in this section. Since LTE supports user speed up to 500 km/hr, Fig. 5.8 shows results with very high user speed. The results show more stable P_{R_Min} in the proposed method.

Figure 5.5: The lowest received power (P_{R_Min}) vs. user speed ($Q_{hyst}SF$ = -0.2 dB).

Figure 5.6: The lowest received power (P_{R_Min}) vs. user speed ($Q_{hyst}SF$ = -0.4 dB).

Figure 5.7: The lowest received power (P_{R_Min}) vs. user speed
($Q_{hyst}SF$ = -0.4 dB, trajectory 2).

Fig. 5.10 shows results with the selection of higher SSSF as the user speed increases. Thus, a set point is established using a fixed $Q_{hyst}SF$, provided by the eNodeB and then the UE adapts SSSF with user speed at its own discretion. Fig. 5.11 shows the cell switching

delay vs. user speed. The delay is calculated as the time that elapses between the moment of crossing the cell boundary and the moment of triggering cell reselection. The delay does not fall too low to pose risk of ping-pong in the proposed method.
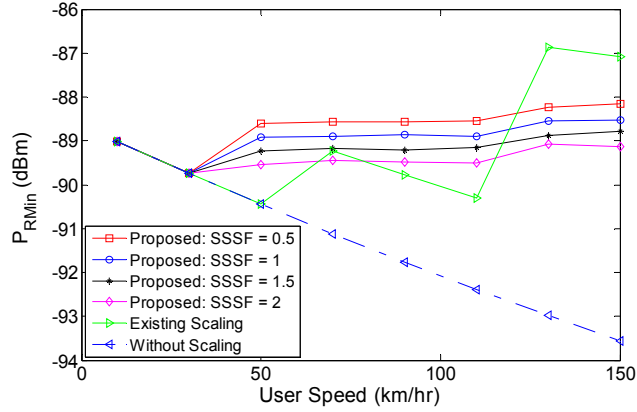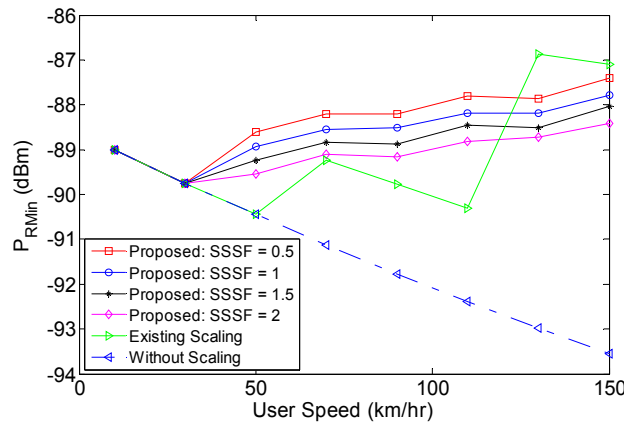

Figure 5.8: The lowest received power (P_{R_Min}) vs. user speed ($Q_{hyst}SF = -0.4$ dB).

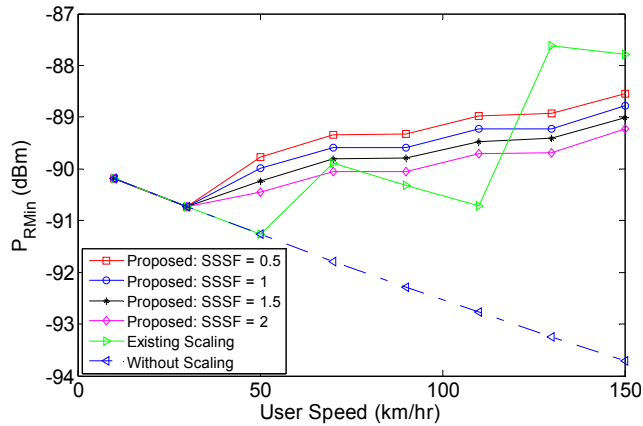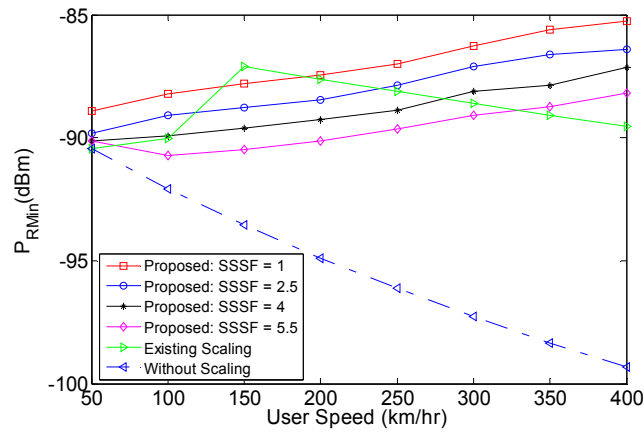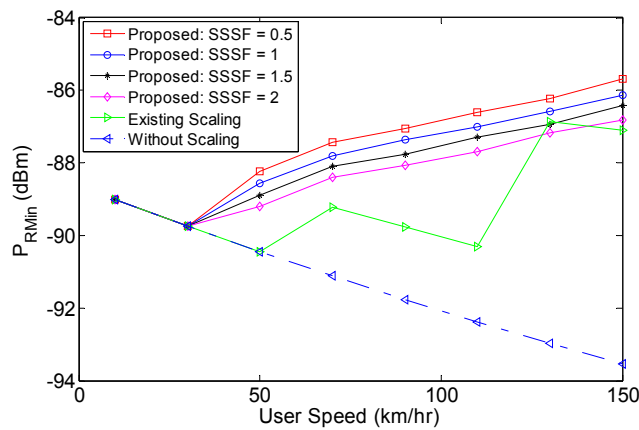**Figure 5.9:** The lowest received power (P_{R_Min}) vs. user speed ($Q_{hyst}SF = -0.8$ dB).
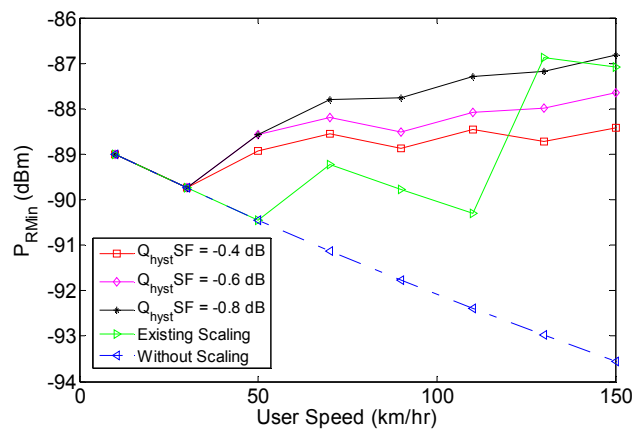
Figure 5.10: The lowest received power (P_{R_Min}) vs. user speed with adaptation of SSSF by the UE for $Q_{hyst}SF = -0.4$ dB, -0.6 dB and -0.8 dB.
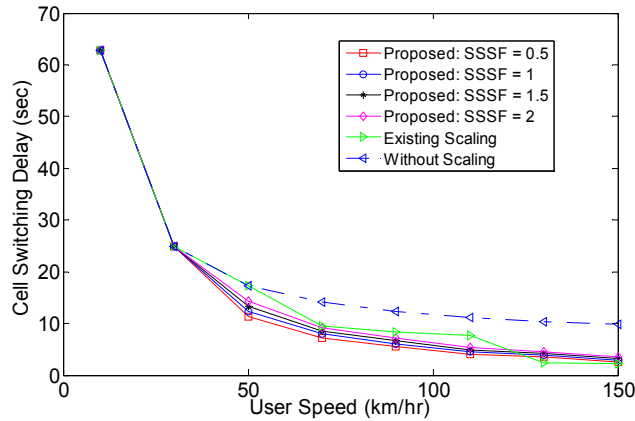
Figure 5.11: Cell switching delay vs. user speed (Q_{hyst}SF = -0.4 dB).

5.5.2   Results for Handover

The UE sends Only trajectory 1 is used for simulation in the case of handover. Fig. 5.12 shows P_{R_Min} vs. user speed profile in the proposed method but the scaling of hysteresis is avoided. Here, P_{R_Min} is significantly better in the proposed method but its stability with the user speed is not a lot better.

Figure 5.12: The lowest received power (P_{R_Min}) vs. user speed avoiding scaling of hysteresis in the proposed method (Hyst-SF= -0.4 dB).

Fig. 5.13 and Fig. 5.14 show P_{R_Min} vs. user speed with scaling of hysteresis in the proposed method for two different values of Hyst-SF. Evidently, P_{R_Min} achieves much better values and stability in the proposed method compared to the existing method. Fig. 5.15 shows the cell switching delay vs. user speed with scaling of hysteresis.

102

Figure 5.13: The lowest received power (P_{R_Min}) vs. user speed with scaling of hysteresis in the proposed method (Hyst-SF= -0.4 dB).

Figure 5.14: The lowest received power (P_{R_Min}) vs. user speed with scaling of hysteresis in the proposed method (Hyst-SF= -0.7 dB).
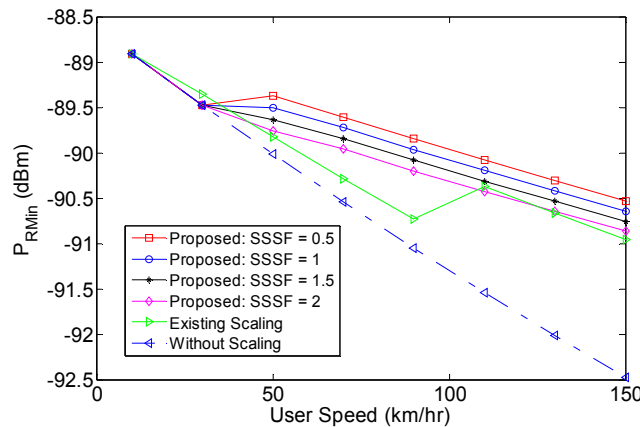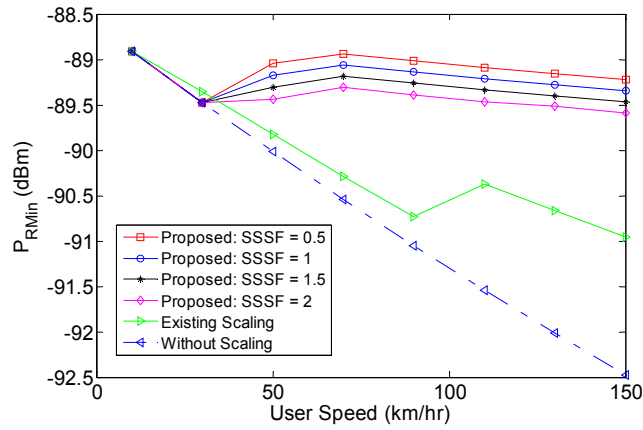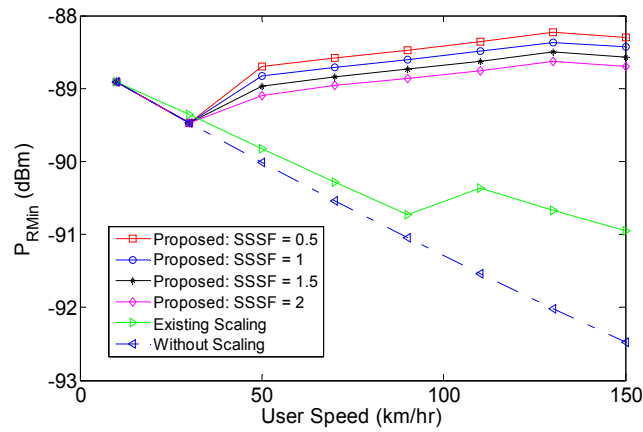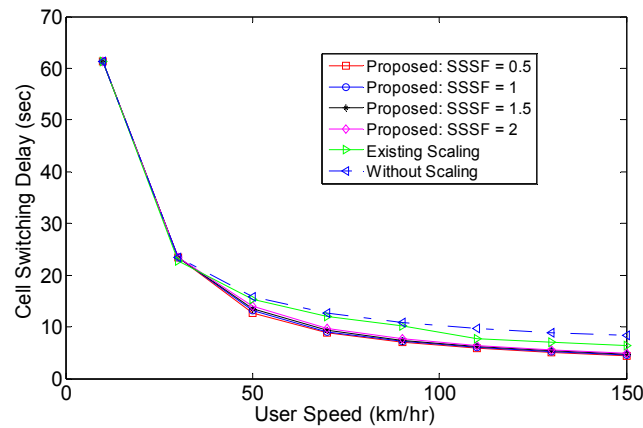

Figure 5.15: Cell switching delay vs. user speed with scaling of hysteresis in the proposed method (Hyst-SF= -0.4 dB).

5.6   Analysis of Results

The MATLAB based simulation is limited to partially implement the proposed features. A strong system level simulator is required for comprehensive validation of the proposals. But such a simulator cannot be made accessible at the moment and so, a proper simulation has been left as a future work. However, it has been logically explained that the proposed method has ample controllability and it can potentially have remarkable achievements.

The proposed and existing methods were simulated and compared. The parameter values in the simulation assumptions comply with 3GPP specifications when certain specified values are available. The same values of α, β and η were used in all simulations because the simplified simulation environment is unable to demonstrate discrepancies with variation in their values.

The simulation results show that the lowest received power P_{R_Min} does not fall with user speed in the proposed method as low as it does in the existing method. The decreases in SSSF can trigger cell switching earlier boosting P_{R_Min}. If the absolute value of $Q_{hyst}SF$ is higher, then it results in higher P_{R_Min} values for the same SSSF values. The proposed method attains higher stability in P_{R_Min} with variation of user speed. The proposed method establishes a set point using a fixed value of $Q_{hyst}SF$ and then varies SSSF values as the user speed varies. Decreases in SSSF can trigger cell switching earlier boosting P_{R_Min}. With such controllability, P_{R_Min} values can be made as velocity independent as seems justified.

The simulation results for handover show that if the scaling of hysteresis is avoided, P_{R_Min} stays significantly higher in the proposed method compared to the existing method but its stability with the user speed is not a lot better. The proposed method applies scaling of hysteresis and with this scaling, P_{R_Min} stays far higher in the proposed method compared to the existing method. The stability of P_{R_Min} is way above in the proposed method compared to the existing method.

In the case of handover, the lowest value of cell switching delay in the simulation results was 4417 msec and it occurred at user velocity 150 km/hr with SSSF 0.5. In a measurement period of 200 ms, the physical layer of the UE can perform measurements of at least 8 identified intra-frequency cells and report the results to layer 3 when no measurement gaps are activated [18]. In 4417 msec, the layer 3 receives measurement results at least 22 times from the physical layer until the layer 3 sends its last measurement report. This appears

to be sufficient to ensure that the neighbor cell has really got better than the serving cell avoiding the possibility of ping-pong effects.

5.7   Discussion and Conclusion

It can be inferred from many different challenges shown in Section 5.2 that a versatile controllability over cell switching is required to address the challenges. So, the proposed scheme uses the following changes in the scaling of cell switching parameters.

i.   An adaptive scaling of cell switching parameters is used and also, a versatile controllability over the scaling is used so that the scaling can better fit into the current status considering user speeds, traffic loads, street patterns and types of cells involved in switching.

ii.   In order to improve versatility in the controllability, instead of using only the recent history of cell reselections or handovers, it is combined with the Doppler spread based velocity estimation for computation of the scaling factor. The rationale behind this action is that one between the number of cell switching events and Doppler spread based velocity estimation can be found better indicative of the requirement of scaling than the other depending on the scenario in HetNets. Secondly, since the scaling with speed reduces the cell switching delay, it increases the possibility of ping-pong effects. The inclusion of Doppler spread ensures that the velocity is really high when high scaling is applied and the chances of ping-pong effects are thus mitigated.

iii. The scaling varies smoothly instead of varying only at two discrete steps. Only normal and high mobility states are used leaving the medium mobility state out. A smooth scaling is performed when the UE is in high mobility state and the normal mobility uses no scaling. Because of the smooth scaling, P_{R_Min} does not fall too low and the cell edge performance does not degrade too much. This may preclude RLF and HF. Also, P_{R_Min} becomes more independent of the user speed. Due to the stability of P_{R_Min}, the receiver can be more easily designed with proper sensitivity. Since the medium mobility state is removed, there is no need to transmit N_{CR_M}.

iv.   Instead of letting eNodeB gain the whole control over scaling, a part of the control is shifted to the UE. The eNodeB controlled scaling of the hysteresis margin is used to establish a set point. Then the UE performs adjustment of timer period as an overlay around the set point at the discretion of the UE itself. This adjustment does not require

any feedback or any overhead and it can quickly adapt to the changes. Thus, it can function as a fine-tuning on top of the eNodeB controlled scaling.

The proposed method provides high flexibility and adaptability in controlling the scaling of cell switching parameters for speedy users in HetNets. It can adapt to the continuously changing conditions of coverage, loading and user speed. It achieves a smoother adaptation of cell switching delay with the user speed. It improves P_{R_Min} values and better stabilizes P_{R_Min}. The possibility of ping-pong effects does not increase and it is rather mitigated. $Q_{hyst}SF$ or Hyst-SF and SSSF allow independent control from both the eNodeB and the UE. However, the proposed method adds computational load and increases signaling overhead to some extent. The implementation of the proposed method seems warranted.

The interference issues are a great concern in HetNets. Some sub-optimal solutions to mitigate interference are popular but there can be investigation for a solution with better performance. Also, the resource share among large and small cells can be coordinated in HetNets for more efficient utilization. Such attempts will be discussed in the next chapter.

Chapter 6

Resource Allocation and Sectorization for FFR in Macro-Femto Based HetNets

The ever increasing demand for higher data rates and better quality of services has motivated the consideration of heterogeneous networks (HetNets) in LTE as discussed in Section 2.7. The femtocells, introduced in Section 2.6, are very popular components of HetNets. In many cases, the HetNets are comprised of only macrocells and femtocells and they can be referred to as macro-femto based HetNets or in short, femtocell systems. Effective distribution of resource allocations and sectorization can improve the performance of femtocell systems significantly and contribute in meeting the growing data rate demand.

6.1   Introduction

The Fractional Frequency Reuse (FFR) is a resource allocation technique that can effectively mitigate inter-cell interference (ICI) in macro-femto based HetNets or femtocell systems. Various FFR schemes have been suggested to address the challenge of interference in femtocell systems. Section 3.5.2 shows that the MAC layer of the eNodeB allocates the available radio resources among different UEs in a cell and it is possible that different parts of the resources are allocated in different parts of the cell. In this chapter, the scopes of interference mitigation and capacity improvement in FFR deployment for macro-femto based HetNets are discussed. The MAC layer is proposed to allocate resources such that the resource share gradually varies with distance from the eNodeB for both macrocells and femtocells in order to attain better utilization of the resources. This is performed effectively using three layers in the cell. The proposal also employs high number sectors in a cell, low interference and good frequency reuse. MATLAB based Monte-Carlo simulations have been performed, which show that the proposed scheme achieves significantly better throughput compared to the existing FFR schemes.

6.2   Common FFR Schemes for Macro-Femto Based HetNets

Like most literature addressing the interference issues in macro-femto based HetNets or femtocell systems, it is assumed that a good number of femtocells are deployed where significant macro coverage already exists. Since femtocells use the same spectrum as do the macrocells, the co-channel interference (CCI) between macrocells and femtocells must be tolerable [69]. Apart from the CCI among neighboring macrocells, there are, in fact, two types of interference concerns in femtocell systems: co-tier interference, which occurs between neighboring femtocells and cross-tier interference, which occurs between femtocells and macrocells. These interferences can arise in six possible scenarios as shown below.

1. Femto DL is affected by macro eNB.
2. Femto UL is affected by macro-users.
3. Macro DL is affected by HeNB.
4. Macro UL is affected by femto-users.
5. Femto DL is affected by another HeNB
6. Femto UL is affected by femto-users.

Scenario #1 can have substantial interference because macro eNB typically uses much higher power than HeNB. Scenarios #2 and #3 can have substantial interference when the macro-users stay close to the HeNB. Scenario #4 offers substantial interference when there are a large number of femtocells deployed. Then the aggregate UL interference from the large number of femto-users becomes high. Scenarios #5 and #6 exhibit substantial interference when the femtocells are randomly located, which is more likely in residential deployments. In public or enterprise deployments, femtocells may be carefully located at predesigned places limiting the interference [70]. Effective ICIC techniques need to be used to successfully mitigate interferences in femtocell systems.

Hard Frequency Reuse (HFR) splits the system bandwidth into a number of distinct sub-bands based on a chosen frequency reuse factor and the neighboring cells transmit on different sub-bands as shown in Fig. 6.1 and Fig. 6.2. The user interference at cell edge is maximally reduced in HFR but the spectrum efficiency drops by a factor equal to the reuse factor. An important type of ICIC technique is Fractional Frequency Reuse (FFR), which is the opposite of HFR. FFR allows the use of orthogonal frequency resources among neighboring cells around the cell borders. FFR allows the UEs near the base station to reuse the same subcarriers in neighboring cells which makes frequency reuse factor, 1. But for UEs

closer to the cell edge, the subcarriers are allocated in a coordinated manner. FFR provides good performance gain with low complexity.
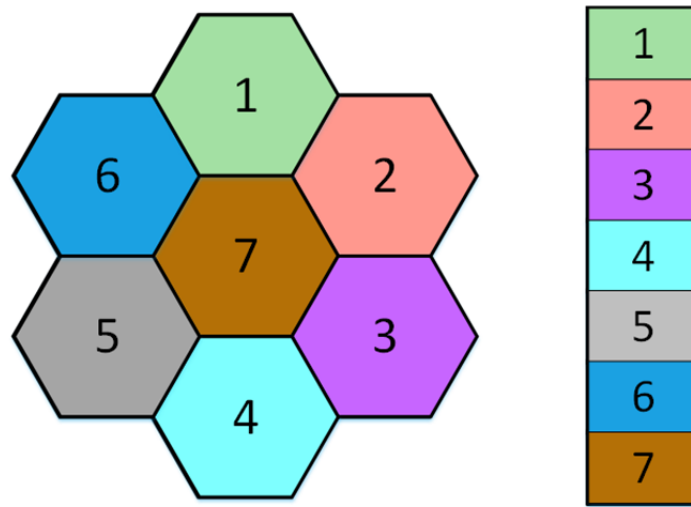
Figure 6.1: Bandwidth allocation in HFR
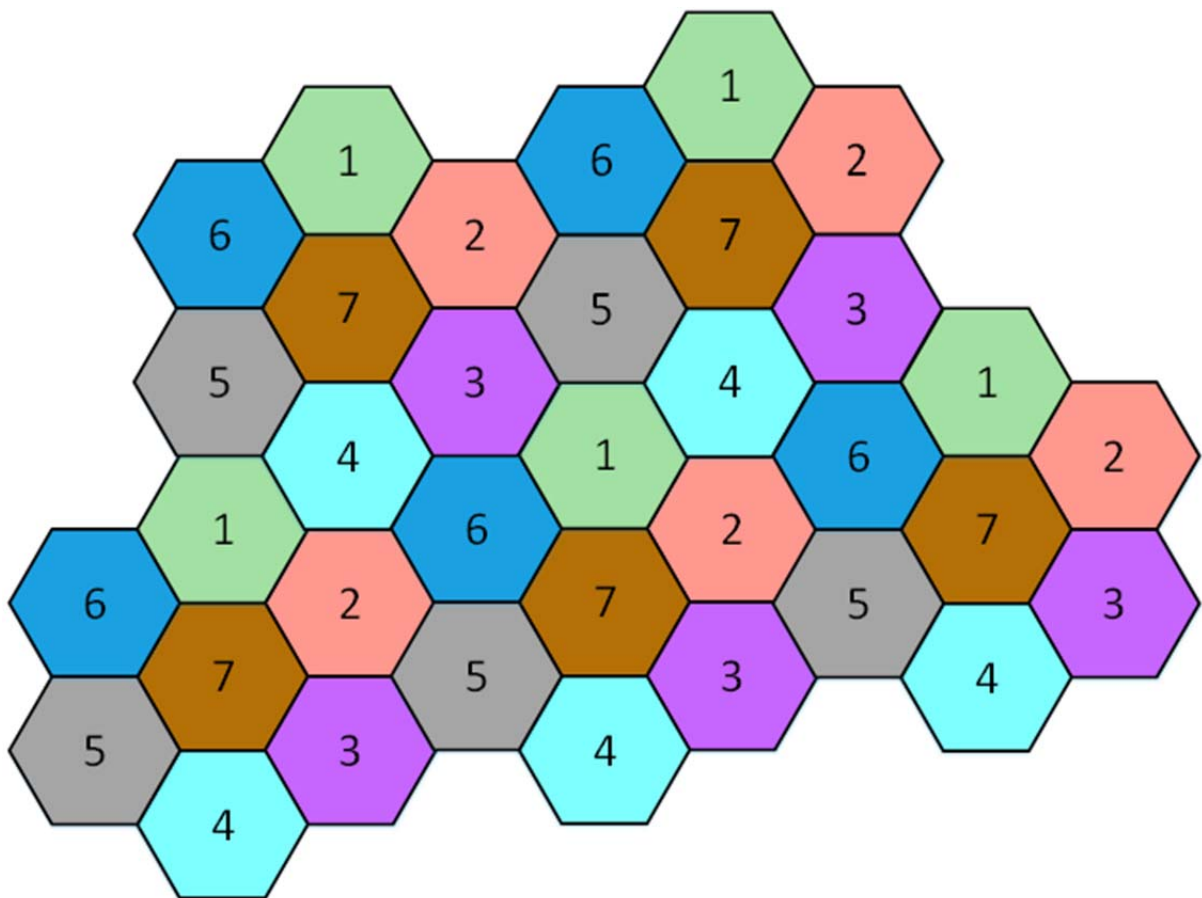
Figure 6.2: HFR deployment showing the neighboring cells

The commonly known schemes of FFR, while considering deployment in femtocell systems, are briefly described below.

6.2.1 Strict FFR

In strict FFR, a cell is partitioned into two regions in this scheme: the center zone and the edge zone. For a cluster of N cells, the frequency band is divided into (N+1) sub-bands as shown in Fig. 6.3 and Fig. 6.4. The center zones of all cells are allocated a common sub-band. The rest of the sub-bands are allocated to the edge zone of the cells depending on the frequency reuse factor (FRF). Here, FRF is 1 and N for the center zone and the edge zone, respectively. A HeNB in the center zone uses the sub-band that is used by the eNB in the edge zone. Similarly, a Home eNB in the edge zone uses the sub-band that is used by the eNB in the center zone. [71] shows that the throughput maximizes when the center zone radius is set to 0.65 times the macrocell radius.

Figure 6.3: Bandwidth allocation in strict FFR

6.2.2 Soft FFR

In soft FFR, a cell is partitioned into center zone and edge zone. For a cluster of N cells, the frequency band is divided into N sub-bands. The edge zones of the cells in a cluster are allocated different sub-bands and the center zone uses the sub-band selected for the neighboring cell's edge zone. Similar to strict FFR, a HeNB in the center zone uses the sub-

110

band that is used by the eNB in the edge zone and a HeNB in the edge zone uses the sub-band that is used by the eNB in the center zone.
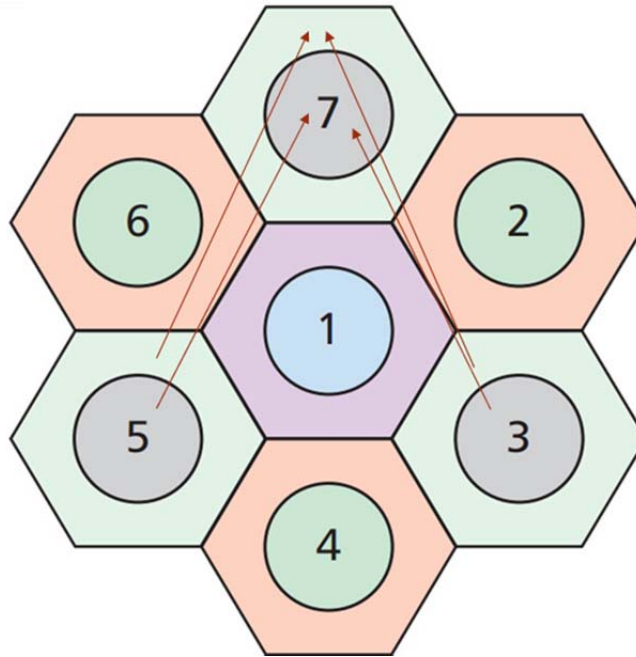
Figure 6.4: Strict FFR deployment showing the neighboring cells

6.2.3   FFR-3

In FFR-3, a cell is partitioned into center zone and edge zone along with three sectors as shown in Fig. 6.5 and Fig. 6.6. The frequency band is divided into four sub-bands, namely, A, B, C and D where sub-band A is made larger than others. Sub-band A is allocated to the center zone and the remaining three sub-bands are allocated to the edge zone of the three sectors. Here, FRF is 1 and 3 for the center zone and the edge zone, respectively. A HeNB in the center zone uses all the sub-bands except the sub-band that is used by the eNB in the center zone and a HeNB in the edge zone uses all the sub-bands except the one that is used by the eNB in that particular zone. [72] shows that the throughput maximizes when the center zone radius is set to 61 percent of the macrocell radius.
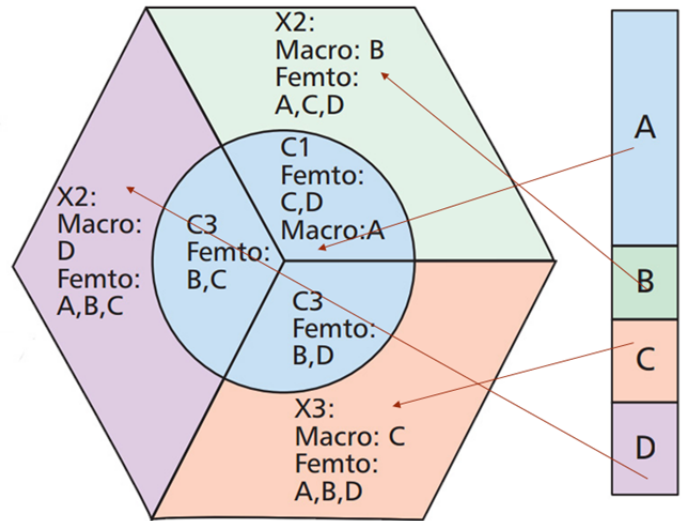
Figure 6.5: Bandwidth allocation in FFR-3
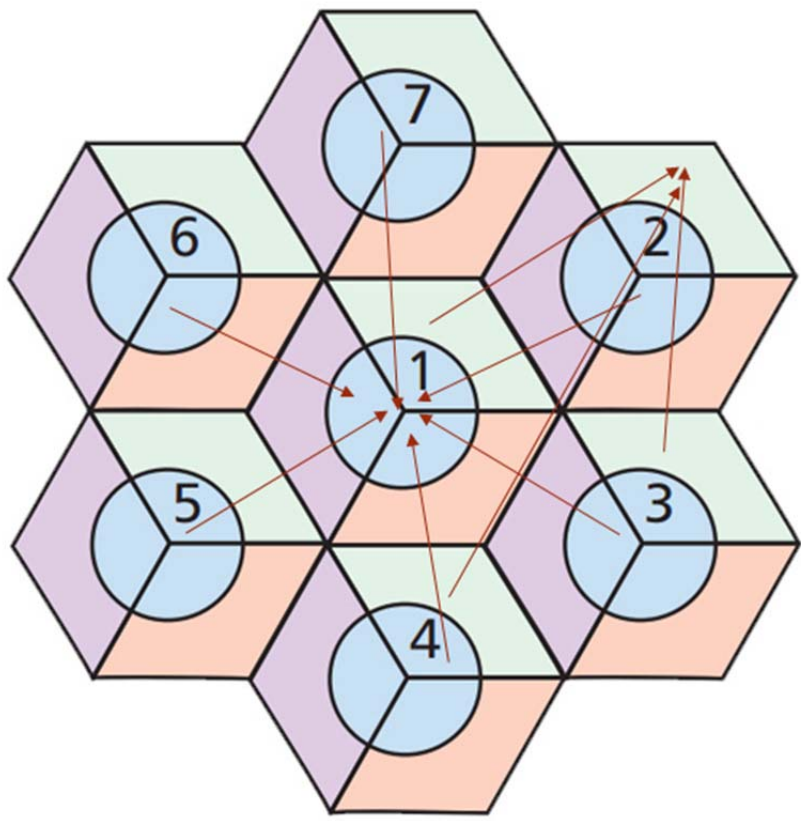
Figure 6.6: FFR-3 deployment showing the neighboring cells

112

6.2.4   Optimal Static FFR (OSFFR)

[72] proposes OSFFR. Here, a cell is partitioned into center zone and edge zone along with six sectors. The frequency band is divided into seven sub-bands where one of the sub-bands is made larger than others and it is allocated to the center zone. The remaining six sub-bands are allocated to the edge zone of the six sectors as shown in Fig. 6.7 and Fig. 6.8. Here, FRF is 1 and 6 for the center zone and the edge zone, respectively. A HeNB in the center zone uses all sub-bands, except the one used in the center zone by the eNB as well as the sub-bands used by the eNB in the edge zone of the same sector and in the two edge zones adjacent to that sector. A HeNB in the edge zone uses all sub-bands, except the sub-band used by the eNB in the same edge zone. [72] shows that the throughput maximizes when the center zone radius is set to 54 percent of the macrocell radius.
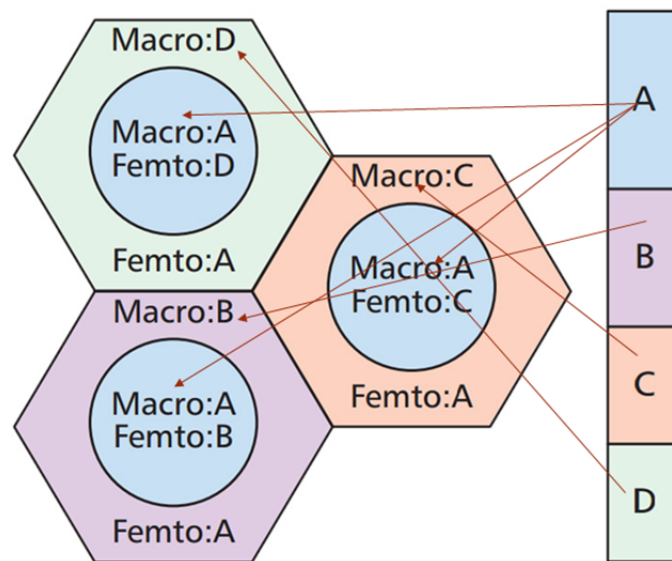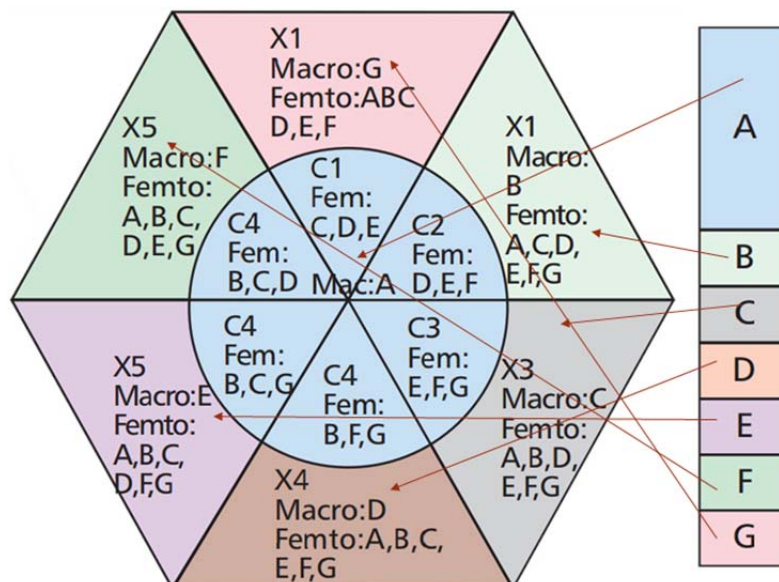

Figure 6.7: Bandwidth allocation in OSFFR

6.2.5   3-Layer/3-Sector FFR

[73] proposes this scheme. Here, a cell is partitioned into three zones: inner or center zone, intermediate zone and outer zone along with three sectors. The frequency band is divided into four sub-bands. One sub-band is allocated to the center zone. The rest three sub-bands are allocated to the intermediate zones with FRF 3. Similarly, these three sub-bands are also allocated to the outer zones with FRF 3. However, the same two sub-bands are not allocated to the intermediate zone and outer zone of the same sector. A HeNB, whether it is

in the center zone, intermediate zone or outer zone, uses all sub-bands except the one used by the eNB in its own zone.
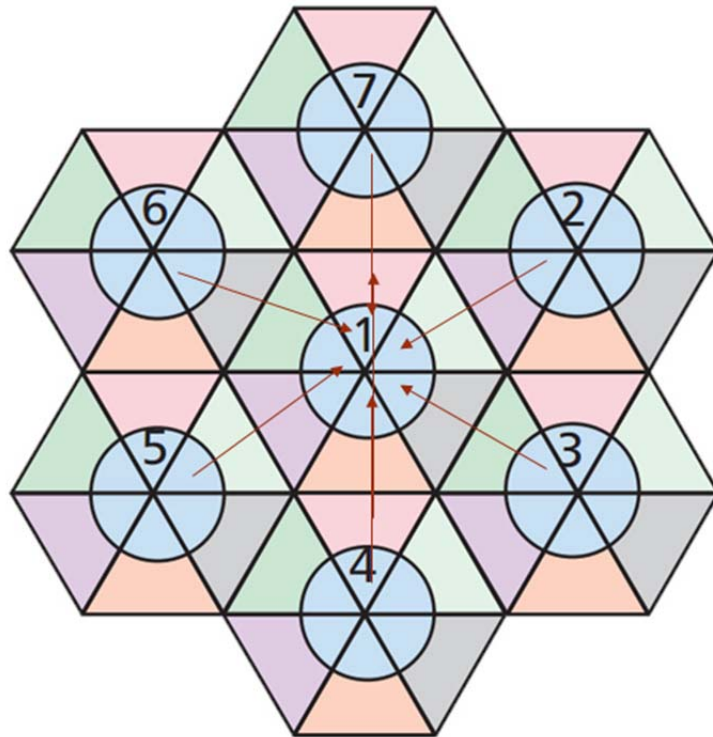
Figure 6.8: OSFFR deployment showing the neighboring cells
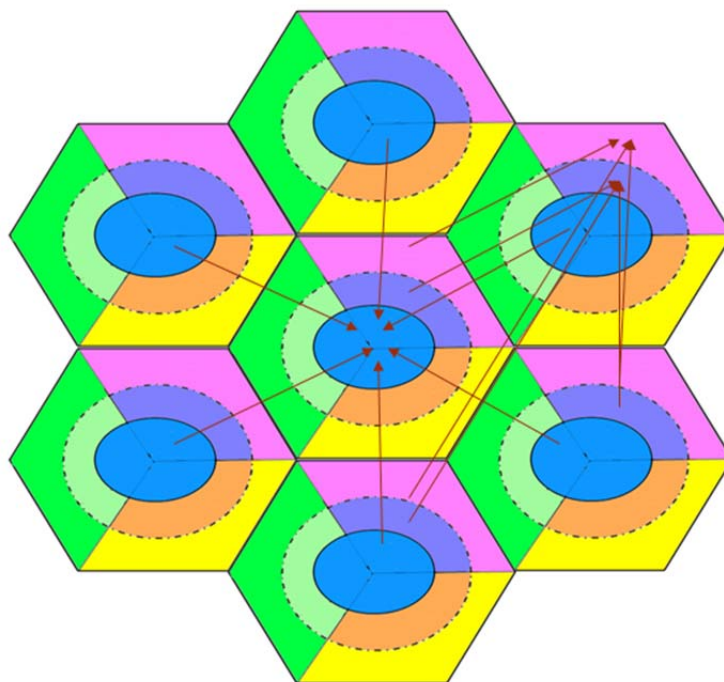
Figure 6.9: 3-Layer/3-Sector FFR deployment showing the neighboring cells

6.3   Related Work

To mitigate the interference in femtocell systems, various ICIC techniques have been suggested. One of the major ICIC techniques is fractional frequency reuse (FFR), which also yields better frequency reuse [74]−[77]. In case of FFR, the cell is divided into different regions and the available frequency band is also divided into several sub-bands. Then different sub-bands are allocated to different regions. Also, different power levels may be used for different sub-bands in different regions. A FFR deployment mitigating interference in femtocell systems successfully can achieve high spectral efficiency. To address the interference issues effectively, a number of FFR deployment schemes have been suggested. The strict FFR [76], soft FFR [78], FFR-3 [75][79], optimal static FFR (OSFFR) [72] and 3-layer/3-sector FFR [73] are examples of such schemes and they employ different methods in splitting the cells, splitting the available bandwidth and allocating different resources to different regions. [80] uses graph theory and [81] uses two-stage heuristic approach to find optimal design of FFR deployment.

It may be noted that the co-tier and cross-tier interferences arise because of overlapping of coverages. These interferences become trivial when, unlike the assumption in this chapter, there are only a few femtocells deployed at a good distance from the macro eNB where the macro coverage is very poor. In this case, the interference management typically does not take up a great challenge and the ordinary interference management techniques, for example, simple FFR schemes or hard frequency reuse (HFR), may be chosen. The capacity improvement scopes for femtocell systems with FFR deployment are investigated. The capacity improvement problem has been formulated as an optimization problem, whose objective is to maximize the throughput with the constraints of transmit power limits. The complexity of the problem calls for sub-optimal solutions. A sub-optimal solution is proposed and simulation results show that the proposed scheme attains higher capacity or throughput compared to the existing schemes.

6.4   System Model and Problem Formulation

Ahexagonal multi-cell LTE femtocell based network is considered with FFR deployment. It is assumed that the sectors in a cell are sequentially numbered clockwise around the cell and represented as s and there are S number of sectors in each cell. In addition, there can be different regions based on the distance from the macro eNB in each

sector. This kind of regions in each sector will be termed as layers. It is assumed that the layers in a sector are represented as l where $l = 1$ represents the center layer. l sequentially increases outwards and there are L number of layers in each sector. It is assumed that the sub-carriers are represented as k and there are K number of sub-carriers in the whole available bandwidth. The downlink SINR for macro-user m on sub-carrier k can be expressed by

$$SINR_m^k = \frac{P_M^k G_{m,M}^k}{N_0 \Delta f + \sum_{M' \neq M} P_{M'}^k G_{m,M'}^k + \sum_F P_F^k G_{m,F}^k} \tag{6.1}$$

where $G_{m,M}^k$ and $G_{m,M'}^k$ represent the channel gain, on sub-carrier k, between macro-user m and serving macrocell M and between user m and neighbor macrocells M', respectively. P_M^k and $P_{M'}^k$ are the transmit power on sub-carrier k, of serving cell M and whole neighbor cell set M', respectively. P_F^k is the transmit power of neighbor femtocells F on sub-carrier k. $G_{m,F}^k$ represents the channel gain between user m and neighbor femtocells on sub-carrier k. N_0 is the white noise power spectral density, and Δf refers to the sub-carrier spacing. Δf is set to 15 kHz in LTE.

Similarly, the downlink SINR for femto-user f on sub-carrier k can be expressed by

$$SINR_f^k = \frac{P_F^k G_{f,F}^k}{N_0 \Delta f + \sum_M P_M^k G_{f,M}^k + \sum_{F' \neq F} P_{F'}^k G_{f,F'}^k} \tag{6.2}$$

where $G_{f,F}^k$ and $G_{f,F'}^k$ represent the channel gain, on sub-carrier k, between femto-user f and serving femtocell F and between user f and neighbor femtocells F', respectively. P_F^k and $P_{F'}^k$ are the transmit power on sub-carrier k, of serving cell F and whole neighbor cell set F', respectively. P_M^k is the transmit power of neighbor macrocells M on sub-carrier k. $G_{f,M}^k$ represents the channel gain between user f and neighbor macrocells on sub-carrier k.

A macro- or femto-user is represented as u where $u \in \{m, f\}$. The channel gain is dominantly affected by Rayleigh fading H, log-normal shadowing X_α, and path losses PL_u. So, the channel gain of a user u on sub-carrier k can be shown as follows [82]:

$$G_{u,M\,or\,F}^k = 10^{(-PL_u + X_\alpha)/10} \left| H_u^k \right|^2 \tag{6.3}$$

Path loss is calculated according to two formulas shown in Table 6.2 [79]. The practical capacity of a user u on sub-carrier k can be given by

$$C_u^k = \Delta f \cdot \log_2 (1 + \alpha SINR_u^k) \tag{6.4}$$

where α a is a constant for target Bit Error Rate (BER), and defined by $\alpha = -1.5 / \ln(5BER)$ [82]. Here, BER is set to 10^{-6}.

It is assumed that there are $N_m^{s,l}$ and N_f^s number of macro- and femto-users in layer l of sector s, respectively. The overall cell throughput for macro-users can be expressed as follows:

$$R_M = \sum_{s=1}^{S} \sum_{l=1}^{L} \sum_{m=1}^{N_m^{s,l}} \sum_{k=1}^{K} \beta_m^k C_m^k \tag{6.5}$$

where β_m^k represents the sub-carrier assignment to macro-users. When the sub-carrier k is assigned to macro user m, $\beta_m^k = 1$ and otherwise, $\beta_m^k = 0$. Thus,

$$\beta_m^k \in \{0,1\}, \forall m. \tag{6.6}$$

Similarly, the overall cell throughput for femto-users can be expressed as follows:

$$R_F = \sum_{s=1}^{S} \sum_{l=1}^{L} \sum_{f=1}^{N_f^{s,l}} \sum_{k=1}^{K} \beta_f^k C_m^k \tag{6.7}$$

where β_f^k represents the sub-carrier assignment to femto-users. When the sub-carrier k is assigned to femto-user f, $\beta_f^k = 1$, and otherwise, $\beta_f^k = 0$. Thus,

$$\beta_f^k \in \{0,1\}, \forall f. \tag{6.8}$$

The total transmit power of a macro eNB P_{eNB} cannot exceed its maximum transmit power capability $P_{eNB}^{\max}$. Therefore,

$$P_{eNB} = \sum_{s=1}^{S} \sum_{l=1}^{L} \sum_{m=1}^{N_m^{s,l}} \sum_{k=1}^{K} \beta_m^k P_M^k \le P_{eNB}^{\max} \tag{6.9}$$

Similarly, the total transmit power of a HeNB P_{HeNB} cannot exceed its maximum transmit power capability $P_{HeNB}^{\max}$. Assuming that a Home eNodeB serves a total of N_f^F number of femto-users,

$$P_{HeNB} = \sum_{f=1}^{N_f^F} \sum_{k=1}^{K} \beta_f^k P_F^k \le P_{HeNB}^{\max} \tag{6.10}$$

The total cell throughput is $R_T = R_M + R_F$. So, the optimization problem can be formulated as follows while satisfying (9) and (10):

$$\max_{S,L,\beta_m^k,\beta_f^k} R_T = \max_{S,L,\beta_m^k,\beta_f^k} \sum_{s=1}^{S} \sum_{l=1}^{L} \sum_{m=1}^{N_m^{s,l}} \sum_{k=1}^{K} \beta_m^k C_m^k$$
$$+ \sum_{s=1}^{S} \sum_{l=1}^{L} \sum_{f=1}^{N_f^{s,l}} \sum_{k=1}^{K} \beta_f^k C_m^k \tag{6.11}$$

The optimization problem is NP-complete for which the solution is not evident. FFR schemes described in Section 6.2 are essentially a few existing or already proposed sub-

optimal solutions. A sub-optimal solution is proposed that achieves higher overall throughput identifying the solution closer to the optimum.

6.5   Proposed Sub-optimal Solution

A greedy heuristic algorithm is proposed for resource allocation by the MAC layer with an attempt for a sub-optimal solution with better throughput [3]. Some considerations are described below that help reach the sub-optimal solution. It is assumed that the assignment of sub-carrier k in layer l of sector s for macro- and femto-users are represented by $\beta_{m,s,l}^k$ and $\beta_{f,s,l}^k$, respectively. $\beta_{m,s,l}^k$ or $\beta_{f,s,l}^k$ are set to 1, if the sub-carrier k is assigned and set to 0, otherwise.

i.   To limit CCI, a sub-carrier can be allocated to only one macro-user in each sector. So,

$$\sum_{l=1}^{L}\sum_{m=1}^{N_m^{s,l}} \beta_m^k \in \{0,1\}, \forall k.$$

ii.   To limit cross-tier interference, it is assumed that the same sub-carriers are not allocated to macro- and femto-users in the same layer in a sector. Thus,

$$\beta_{m,s,l}^k \neq \beta_{f,s,l}^k \ \text{ if } \ \beta_{m,s,l}^k = 1, \forall k.$$

iii. The central layer can be allocated sub-carriers using FFR=1. Identifying these sub-carriers as k_C and the center layer as $l = C$,

$$\beta_{m,s,l=C}^k = 1, \text{ and } \beta_{m,s,l \neq C}^k = 0 \ \ , \ \ \ k = k_C.$$

Since 1 is the lowest frequency reuse factor, k_C may include a good number of sub-carriers from the available bandwidth.

iv.   When $k \neq k_C$, to limit CCI, it is assumed that the same sub-carriers are not allocated to macro-users in adjacent sectors in a cell. Since sector $s+1$ and s-1 are adjacent to sector s,

$$\beta_{m,s,l}^k \neq \beta_{m,s+1,l}^k, \text{ if } \ \beta_{m,s,l}^k = 1, \ \ \ \ k \neq k_C$$

and

$$\beta_{m,s,l}^k \neq \beta_{m,s-1,l}^k, \text{ if } \beta_{m,s,l}^k = 1, \ \ \ \ k \neq k_C.$$

v.   Since the two outer layers of adjacent cells lie next to each other, to limit CCI, it is assumed that the same sub-carriers are not allocated to macro-users in these outer layers. In a hexagonal multi-cell network, the sector of the adjacent cell, which is

aligned with sector *s*, can be equivalently found as $\lfloor s + S/2 \rfloor$. Identifying the outer layers as $l = O$,

$$\beta^k_{m,s,l=0} \neq \beta^k_{m,\lfloor s+\frac{S}{2}\rfloor,l=0}, \text{ if } \beta^k_{m,s,l=0} = 1, \ k \neq k_C$$

vi.   When $\beta^k_{f,s,l} = 1$, as *l* decreases, the associated P^k_M can be decreased due to smaller path loss and the sub-carrier allocation causes less interference to neighboring cells. Similarly, as *l* decreases, the associated $G^k_{m,M'}$ decreases due to larger path loss from the neighbor macrocells and the sub-carrier allocation suffers from less interference.

vii.   When $\beta^k_{m,s,l} = 1$, the associated $SINR^k_m$ improves as *l* decreases and so, the macro service can use the sub-carrier more efficiently. On the other hand, when $\beta^k_{f,s,l} = 1$, the associated $SINR^k_f$ improves as *l* increases because of reduced cross-tier interference and so, the femtocell operation can use the sub-carrier more efficiently.

viii. $\beta^k_{f,s,l} = 1$ may be used repeatedly for different femtocells together leading to increased femtocell throughput. However, in this case, the HeNBs need to be installed at predesigned places to limit the co-tier interference and the cross-tier interference as well.

ix.   A macro- or femto-user can only be allocated one or more whole resource blocks (RBs). Each RB consists of 12 sub-carriers and so, it is 180 kHz wide. Let us assume that the *n* numbered RB contains sub-carriers {*n1, n2, n3, .., nx, ..., n12*}. Thus,

if $\beta^k_{m,s,l} = 1$ where $k \in nx$, , then $\beta^{nx}_{m,s,l} = 1, \forall nx.$;

if $\beta^k_{f,s,l} = 1$ where $k \in nx$, then $\beta^{nx}_{f,s,l} = 1, \forall nx.$.

x.   It is possible to vary the resource share more gradually with distance from the eNB using higher value of *L*. *L* = 3 is physically realizable with tight radio planning and [73] proposes three layers in each sector. However, *L* > 3 is formidable and so, it may not be considered.

xi.   Higher sectorization uses narrower antenna beamwidth at the eNB and thus, causes less inter-cell interference. So, a high value of *S* can reduce the inter-cell cross-tier interference. *S* = 6 is physically realizable with tight radio planning and [72] considers six sectors in each cell for FFR. However, *S* > 6 is formidable and so, it may not be considered.

For simplicity, a uniform distribution of the users is assumed. To attain efficient use of sub-carriers, higher resource share is used for macro service in areas closer to the macro eNB and also, higher resource share is used for femto service in areas further away from the macro eNB. In order to materialize such variation in frequency allocation more gradually, a high value for L is used and so, L = 3. [73] already proposes three layers with three sectors in each cell. Besides, to reduce the inter-cell cross-tier interference, the proposal uses six sectors. It is shown that a 6-sector cell, in general, can attain significantly higher cell throughput compared to a 3-sector cell [83]. The use of S = 6 along with L = 3 requires high number of antennas and a lot of cabling [84]. Also, it will engender high number of interferers and overlapping regions and so, a very delicate radio planning will be required [83]. However, these difficulties may be accepted due to the improvement in throughput.

Algorithm 1: Sub-band Allocation Algorithm

1: Let SB_C be the sub-band allocated to the central layer for macro services using FFR=1

2: Make sub-band SB_C larger compared to all other sub-bands

3:    **for** i = 1 to 6 **do**

4:         $O_i^m \leftarrow SB_i$

5:         $I_i^m \leftarrow SB_{i+2} + SB_{i+4}$

6:         $C_i^m \leftarrow SB_C + SB_{i+2} + SB_{i+4}$

7:         $O_i^f \leftarrow SB_C + SB_{i+1} + SB_{i+2} + SB_{i+4} + SB_{i+5}$

8:         $I_i^f \leftarrow SB_C + SB_{i+1} + SB_{i+3} + SB_{i+5}$

9:         $C_i^f \leftarrow SB_{i+1} + SB_{i+3} + SB_{i+5}$

10:   **end for**

Identifying the intermediate layers as $l=I$, $l \in \{O,I,C\}$. There are a total of 18 sub-areas, namely, O1, O2, O3, O4, O5 and O6 in the outer region; I1, I2, I3, I4, I5 and I6 in the intermediate region; C1, C2, C3, C4, C5 and C6 in the center region, as shown in Fig. 6.10. The outer, intermediate and center regions are identified as O_i^m, I_i^m and C_i^m, respectively, for macro services in sector i. Similarly, the outer, intermediate and center regions are identified as O_i^f, I_i^f and C_i^f, respectively, for femto services in sector *i*. The MAC layer applies a resource allocation scheme based on the above considerations with an attempt to minimize all kinds of interferences. To exemplify the resource allocation scheme, the available bandwidth is divided into 7 sub-bands (SBs), denoted as, A, B, C, D, E, F and G. Each sub-band consists of a number of complete resource blocks (RBs). The sub-bands A to F are sequentially numbered twice from 1 through 6 and also, from 7 through 12 (i.e., both SB2 and SB8

represent sub-band B). The sub-band G is denoted as SBC. Algorithm 1 shows an example of allocation of sub-bands in different sub-areas for macro and femto services. Fig. 6.10 and 6.11 illustrate this allocation. In reality, the sub-band sizes may vary and a careful selection of sub-band sizes should be made considering potential data rate requirements.

Table 6.1 demonstrates the relative frequency resource share for macro use as well as femto use in different layers. It is assumed that each of the sub-bands from A to F contains 6 RBs and the sub-band G contains 14 RBs. The radius of center and intermediate layers are chosen 0.4 and 0.7 times the macrocell radius, respectively. For macro services, when the same sub-band is used in adjacent layers, the sub-band is assumed to be uniformly used over the whole area and the frequency resource used per unit area is computed to determine the resource share in each layer. The cell is considered circular in this computation. For femto services, the relative amount of sub-band allocation in different layers is used to compute the relative resource share. As Table 6.1 shows, there is an achievement of ample gradual decrease and increase in resource share with distance from the eNB for macro use and femto use, respectively, leading to an efficient utilization of the resources.
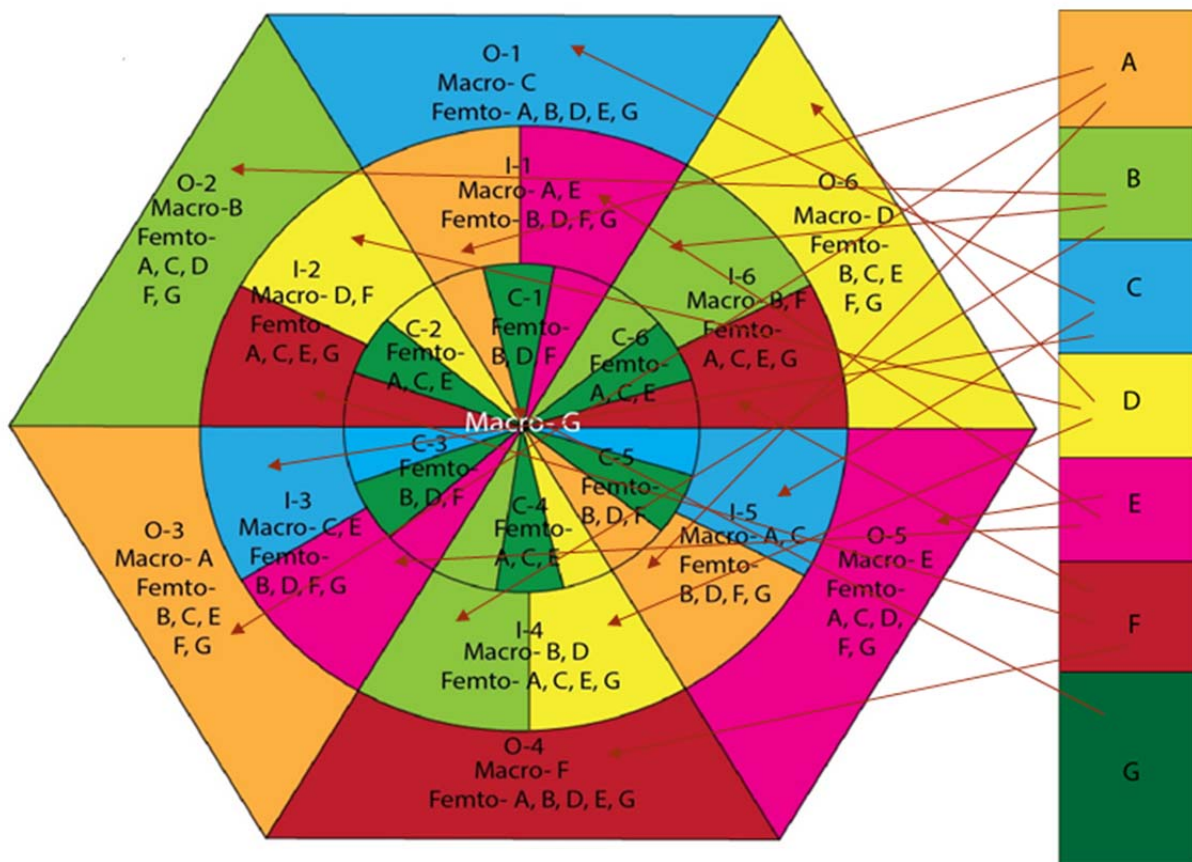

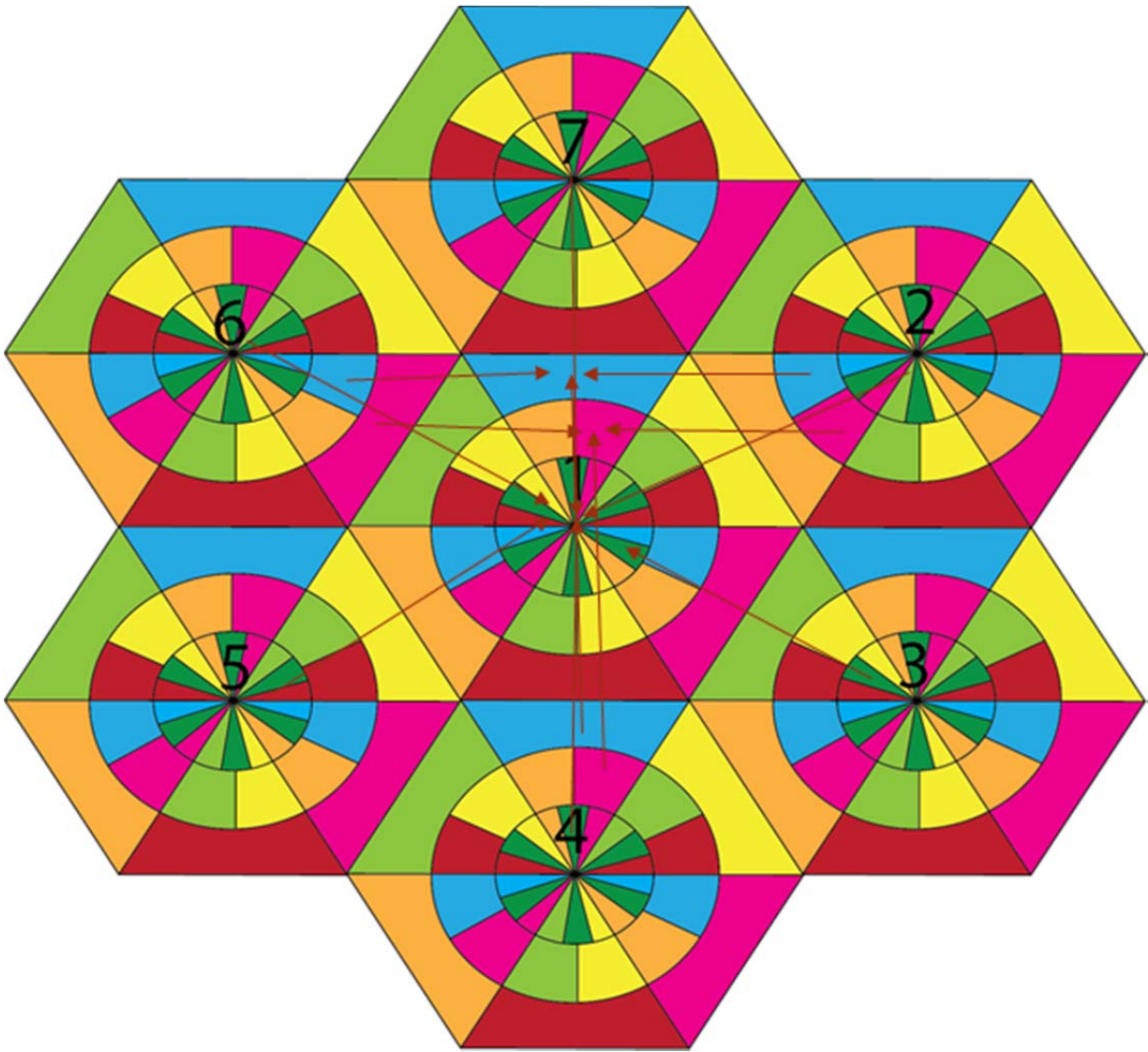**Figure 6.10:** Proposed resource allocation and sectorization scheme

Figure 6.11: Proposed scheme showing the neighboring cells

Table 6.1 Relative resource share in different layers

Layer	Relative Resource Share	
	Macrocell	**Femtocell**
Center	3.321	0.474
Intermediate	2.082	0.842
Outer	1	1

6.6   Simulation

The proposed method was simulated using parameter values shown in Table 6.2. A MATLAB based Monte Carlo simulation is used with one thousand iterations allowing random locations for both UEs and femtocells. It considers a one-tier LTE network. Both the users and the femtocells are distributed randomly in each layer of each sector of the macrocells. The number of femtocells was varied while evaluating the throughput using equations (6.1) − (6.7). Simulation was also performed for existing schemes, namely, FFR-3, OSFFR and 3-layer/3-sector FFR and these schemes were compared with the proposed scheme. Fig. 6.12 shows that the macro throughput of the proposed scheme is significantly higher than the existing schemes. As the number of femtocell is increased, the macro throughput tends to decrease because the macro-users receive higher interference from the femtocells.

Table 6.2 Simulation assumptions for FFR scheme in HetNets

| Parameter	Macrocell	Femtocell
Number of cells	7	66
Cell radius	280 m	30 m
eNB Transmit Power	22 W, 19 W, 15 W	20 mW
Number of users in a cell	50	1
Channel Bandwidth	10 MHz	
No. of RBs	50	
Sub-band size (A − F)	1.08 MHz (6 RBs)	
Sub-band size (G)	2.52 MHz (14 RBs)	
White noise power density	-174 dBm/Hz	
Canter layer radius	0.4 of macrocell radius	
Intermediate layer radius	0.7 of macrocell radius	
Channel model: path loss (outdoor)	28+35log10(d) dB	
Channel model : path loss (indoor)	PL=38.5+20log10(d)+7 dB, $0 < d \leq 10$ PL=38.5+20log10(d)+10 dB, $10 < d \leq 20$ PL=38.5+20log10(d)+15 dB, $20 < d \leq 30$	
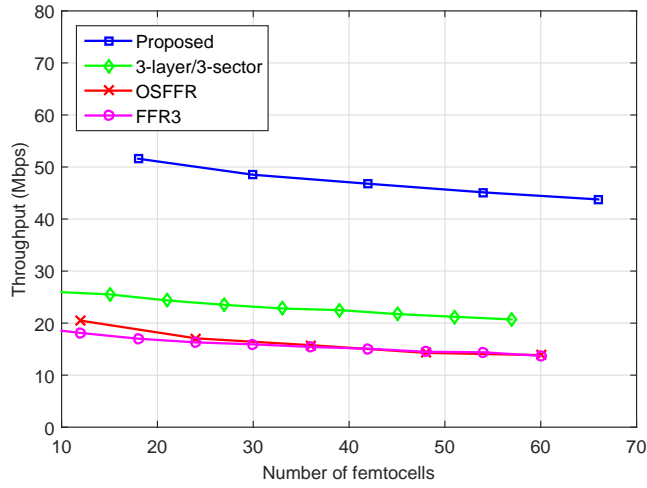
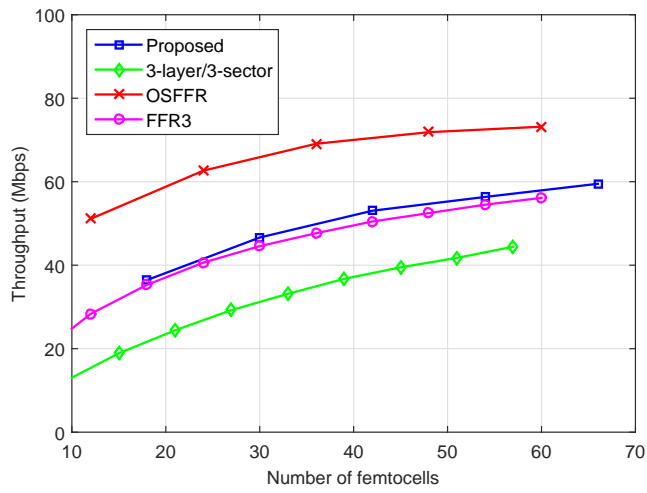**Figure 6.12:** Macro throughput vs. number of femtocells

Figure 6.13: Femto throughput vs. number of femtocells.
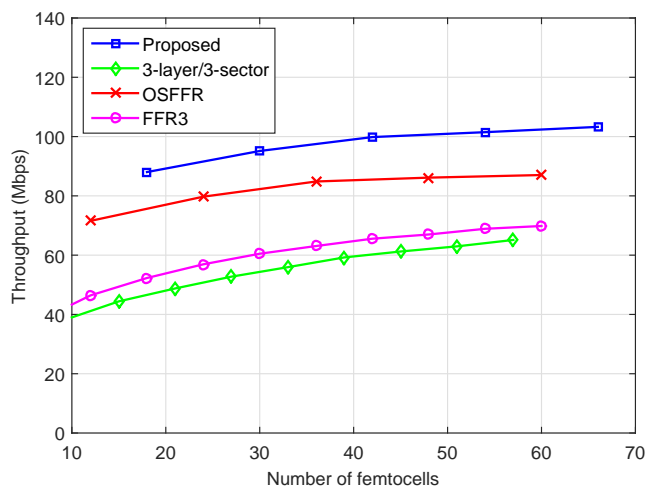
Figure 6.14: Overall throughput vs. number of femtocells.

Fig. 6.13 shows that the femtocell throughput increases with the number of femtocells. However, the growth of the femtocell throughput is fairly impaired primarily because of the co-tier interference and so, the impairment increases with the number of femtocells. The femtocell throughput of the proposed scheme is little higher than FFR-3 and 3-layer/3-sector FFR but it is lower than OSFFR. OSFFR offers much better femtocell throughput because OSFFR, with its two layers, yields less interference to the femtocells. Fig. 6.14 shows the comparison of the overall throughput, which is the summation of throughput of macro and femto cells. Evidently, the overall throughput of the proposed scheme is considerably higher than the existing schemes. Fig. 6.15 shows the macro throughput only in outer and intermediate layers for the proposed scheme and 3-layer/3-sector FFR to demonstrate the individual improvement in those layers.
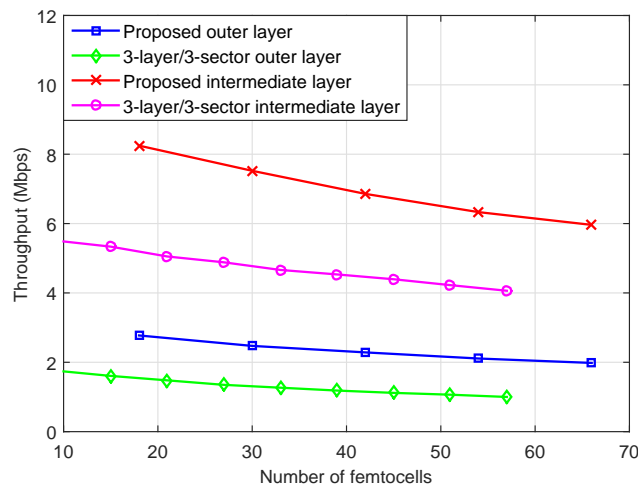
Figure 6.15: Macro throughput in outer and intermediate layers vs. number of femtocells.

6.7 Analysis of Results

The simulation results in Section 6.6 show that the macro throughput and total throughput in the proposed scheme is significantly higher than the existing schemes. The key reasons for this great improvement are the mitigation of interference and efficient utilization of resources. In the case of macro users, the mitigation of interference primarily occurs in the proposed method as follows.

    i.   The high sectorization allows narrow antenna radiation pattern and so, both co-tier interferences and cross-tier interferences will be less in the neighboring cells.

ii. The outer layer uses a part of the resources, which is reused in intermediate and center layers but not reused in outer layers of the nearby cells. This causes less co-tier interference.

iii. The center layer uses a part of the resources, which is not reused in other layers but reused in center layers of the nearby cells. Thus, this part is exposed to very little co-tier interference.

iv. Different parts of the resources are allocated for macro and femto use in the same region. This causes less cross-tier interference. However, there can be cross-tier interference from nearby regions. Therefore, the cross-tier interference increases with the number of femtocells. Consequently, the macro throughput decreases with an increase in the number of femtocells in the simulation results.

As the distance of a macro user increases from the macro eNodeB, the SINR drops because of decrease in signal power and increase in interference. Thus, the resources can be utilized less efficiently. Conversely, the center layer can use the resources very efficiently with a high SINR. Therefore, the resource share is gradually reduced for macro use in the proposed method with distance from the eNodeB. Also, a large dedicated part of the resources have been allocated to the center layer. This gradual variation could be made very well due to the use of three layers.

In the case of femtocells, cross-tier interference was mitigated by allocating different parts of the resources for macro and femto use in the proposed method. The co-tier interference becomes substantial as the number of femtocells grows. The total femtocell throughput grows with the number of femtocells as expected. But the co-tier interference becomes substantial as the number of femtocells grows very high. Thus, the femtocell throughput no longer grows at a good rate when the number of femtocells is too high.

The proposed method also offers efficient utilization of resources for femto users. As the distance of a femto user from the macro eNodeB increases, the SINR improves because of decrease in interference. Thus, the resources can be utilized more efficiently. Therefore, the resource share is gradually increased with distance from the eNodeB in the proposed method. This gradual variation could also be made very well due to the use of three layers.

The femtocell throughput of the proposed scheme is little worse than OSFFR. This is because OSFFR uses only two layers and thus, it allows less cross-tier interference for the femtocells. The overall throughput is the summation of throughput of macro and femto cells.

Primarily because of the great improvement in the macro throughput, the overall throughput of the proposed scheme is significantly higher than all the existing schemes.

6.8　　Discussion and Conclusion

The MAC layer allocates resources to the users. It can identify the UE's position in the cell from the feedback from the UE. So, it is possible that the MAC layer allocates particular parts of the bandwidth to users in particular regions of the cell. This technique for resource allocation is suggested so as to improve performance in HetNets.

The interference is a great concern in HetNets and there are various sub-optimal solutions used as ICIC techniques. Various FFR schemes are regarded as effective sub-optimal solutions as such. This chapter discusses the scopes of interference mitigation and capacity improvement for FFR schemes in HetNets. A resource allocation scheme is proposed, in which the MAC layer uses a gradual variation in resource share for macrocells and femtocells with distance in order to attain effective utilization of the resources. It also attains satisfactory interference mitigation applying high number of layering and sectoring. Since three and six can be regarded as the maximum possible numbers for layers and sectors in a cell, these numbers have been used in the proposal. The proposal also employs good frequency reuse. Monte-Carlo simulations are performed, which show that the proposed scheme is more attractive than the existing schemes in terms of the throughput or the spectral efficiency or the capacity of the system. It achieves throughput significantly higher than the existing schemes.

The femtocell throughput of the proposed scheme is little lower than OSFFR. But the proposed scheme attains much higher macro throughput compared to the existing schemes. In general, the macro service is considered more important compared to the femto service and also, the macro service is usually the major source of revenue for the operator. So, the achievement of the proposed scheme in macro throughput can be greatly helpful.

The proposed scheme requires high number of antennas, a lot of cabling and a very delicate radio planning. But these downsides can be overshadowed by the fact that the achievement of high throughput or spectral efficiency is of paramount importance to meet the present and future data rate demand.

Several parts of the research work have been presented by placing them in a few chapters including this chapter. The whole work is concluded in the next chapter.

Chapter 7

Conclusion

Various aspects of access strata (AS) of LTE have been investigated. Some ideas have been presented with a view to improve the control of AS. One of these aspects is the possible improvement in DRX power saving while there is a tradeoff between power saving and packet delay. An optimum DRX configuration, used by MAC layer, depends on the current data traffic, which is not easy to estimate accurately, especially, for NRT applications. Since NRT applications are delay tolerant, to improve power saving, a DRX method is proposed for use when only NRT applications are running with no RT applications active. The proposed method avoids the estimation of current traffic.

RT traffic is error-tolerant but delay-sensitive. Conversely, NRT traffic is error-sensitive but delay-tolerant. Thus, typically, NRT applications use RLC AM whereas RT applications use RLC UM. The RLC AM retransmits the missing data. The overall delay in the correction of data can be estimated as the summation of the reordering timer value and RTT. The RTT includes round trip propagation delay and processing delay at both ends. Thus, the RLC AM incur some delay and this occurs for NRT traffic. Since the NRT traffic permits a limited amount of delay, a part of this additional delay, permissible for NRT traffic in comparison with RT traffic, is consumed due to RLC retransmissions. After this consumption, some delay is left over, which can be allowed to NRT traffic. The proposed method gradually increases the DRX cycle length to exploit this delay and attains a significant improvement of power saving. Thus, a particular fashion of variation in the DRX cycle length is proposed avoiding continuous estimation of the data traffic when only nonreal-time applications are running with no real-time applications active. The RRC layer configures necessary DRX parameters. The proposal also improves the delay in service resumption after a long inactivity. A stochastic analysis is used assuming M/G/1 queue to validate the improvement of the proposal. The numerical examples exhibit significant power saving in the proposed scheme. A new factor has been defined for a more practical estimation of power saving.

There are challenges to support seamless mobility in HetNets, especially when the user speed is high. These challenges have been investigated. It can be inferred from the

investigation that the cell edge performance is at risk and versatile and rigorous controllability over triggering the cell switching can be beneficial. There should be versatile and rigorous controllability over adaptation of cell switching parameters that simultaneously depends on variation in user speeds, traffic loads, street patterns, types of cells involved in switching, etc. The current proposals related to the adjustment of cell switching parameters lack such controllability. The RRC configures scaling affecting the time to trigger cell reselection and handover depending on the speed of the users. The present method allows varying the scaling only in a few steps and it does not consider the special requirements in the case of various scenarios in HetNets.

A scheme is proposed to scale cell switching parameters that incorporates Doppler spread based velocity estimation and adapts smoothly to various changes. The proposed method employs versatile controllability over the scaling from both the eNodeB and the UE. The scaling is proposed to vary with more granularities and with better control. The proposed scheme can satisfactorily stabilize the experience of minimum signal quality while restricting the chances of ping-pong effects. It enhances flexibility and adaptability in how RRC layer controls scaling leading to optimized scaling based on the HetNet scenario. However, the proposed scheme adds computational load and increases signaling overhead on signaling messages to some extent.

In the case of FFR, the MAC layer employs a resource allocation technique that can effectively mitigate interferences in macro-femto based HetNets. Various FFR schemes are popularly known to address the challenge of interference in femtocell systems. The scopes of interference mitigation and capacity improvement using FFR in HetNets have been investigated. A resource allocation scheme is proposed, in which the MAC layer gradually varies frequency resource share with distance from the eNodeB for both macrocells and femtocells in order to attain better utilization of the resources. This is performed effectively using three layers in the cell. The proposal also employs high number sectors in a cell and it achieves low interference and good frequency reuse. Monte-Carlo simulations are performed, which show that the proposed scheme achieves significantly better throughput compared to the existing FFR schemes. However, the proposed scheme requires high number of antennas, a lot of cabling and a very delicate radio planning.

The MATLAB based simulation results exhibit useful achievements in all the proposed methods. The achievements have been corroborated by logical explanations.

7.1   Potentialities in Real Deployment

The proposed schemes could not be tested either in trial LTE network or using a system level simulator. However, the MATLAB based simulations were performed carefully and the logical explanations were used to establish the potential outcome of the proposed schemes. It is expected that the proposed schemes would be used in real LTE deployments and there would be improvements in performances similar with what are reflected in the simulation results. Thus, the materialization of the proposed methods is a vision.

The proposed schemes often add complexity and signaling overhead and also, they increase costs to some extent. For example, in the proposed FFR scheme, high number of antennas, a lot of cabling and a very delicate radio planning are required. But all these downsides are overshadowed by the achievements of the proposed schemes considering that it is of paramount importance to meet the present and future demand for data rate, power saving and other performances.

The femto throughput in the proposed FFR scheme becomes better than most other current schemes but it is not as good as it is in the case of macro throughput. The achievement in the improvement in macro throughput is remarkable. In general, the macro service is considered more important compared to the femto service and also, the macro service is usually the major source of revenue for the operator. So, the achievement of the proposed scheme in macro throughput can be greatly helpful.

The proposed FFR scheme could further improve its performance if it was possible to transfer some macro-users to femto service in cell edge areas. Therefore, the operation of HeNBs in open access mode can be motivated in outer layers and there, a large number of HeNBs can be installed. This can allow some cell edge users to be served by HeNBs instead of the macro eNB and receive better throughput provided that they are located close to any HeNB.

7.2   Future Work

The proposed schemes presented should be properly validated preferably in trial LTE network or at least using a strong system level simulator. Unfortunately, none of them could be made accessible. So, MATLAB based simulations have been performed instead and the simulation environments are limited in capacities in various ways. Although the logical explanations corroborate the strength and potentialities of the proposed schemes, proper

validation is still required for better establishment of the efficacy of the proposals. Therefore, it is an important future work to manage a capable system level simulator and use it for comprehensive simulation of the proposed schemes. Then testing in a trial LTE network should be performed.

The DRX scheme proposed in Section 4.5 is analyzed using traditional Poisson traffic model for data traffic. However, the traditional Poisson traffic model is not fully accurate when used in LTE data traffic analysis. The European Telecommunications Standards Institute (ETSI) packet traffic model has been recently proposed and this model, with its long range dependence, would be a more suitable choice. It requires getting well versed in this new model in order to use it for such a long analysis. So, formulation of the proposed method using ETSI model can be an important future work.

The proposed scaling of cell switching parameters shown in Section 5.4 uses versatile techniques for high speed users. Some of these techniques can be applied even for low speed users and this should be investigated in future.

The FFR scheme proposed in Section 6.5 limits itself to a fixed resource allocation based on the available radio network planning (RNP) related information. However, an extension of the proposal can be adaptive allocation considering the channel quality of the users at different frequencies as well as their current data rate demands. In this case, the resource share is adjusted with time among eNodeBs, sectors and layers. The channel-dependent scheduling (CDS) will be more efficient with the adaptive allocation. The eNodeBs may use a distributive control of adaptive resource allocation with an attempt for local optimization. For this purpose, new fields can be added to the Load Information message, which currently exists in the specification as a layer 3 message for exchange of information between the RRC layers of neighboring eNodeBs over the X2 interface. These fields may contain information like the estimated interference at different frequencies, channel quality of the users at different frequencies and current load demand in each layer of each sector of the macrocells. It is possible to support the additional overhead in load information message because X2 can be a high speed link.

References

[1] Mohammad T. Kawser, "LTE Air Interface Protocols", Artech House, ISBN: 978-1-60807-201-9, 2011.

[2] Mohammad T. Kawser, Mohammad R. Islam, Ziaul Islam, Mohammad A. Islam, Mohammad M. Hassan, Zobayer Ahmed, Rafid Hasan, "Improvement in DRX Power Saving for Nonreal-Time Traffic in LTE", *ETRI Journal,* 2016.

[3] Mohammad T. Kawser, Mohammad R. Islam, Kazi I. Ahmed, Mir R. Karim, and Jaeed B. Saif, "Efficient Resource Allocation and Sectorization for Fractional Frequency Reuse (FFR) in LTE Femtocell Systems", *Radioengineering Journal*, Vol. 24, No. 4, December, 2015.

[4] Mohammad T. Kawser, Nafiz Imtiaz Bin Hamid, M. Nayeemul Hasan, M. Shah Alam, and M. Musfiqur Rahman, "Downlink SNR to CQI Mapping for Different Multiple Antenna Techniques in LTE", *International Journal of Information and Electronics Engineering,* vol. 2, no. 5, pp. 756-760, ISSN: 2010-3719, 2012.

[5] Mohammad T. Kawser, Hasib M. A. B. Farid, Abduhu R. Hasin, Adil M. J. Sadik, and Ibrahim K. Razu, "Performance Comparison between Round Robin and Proportional Fair Scheduling Methods for LTE", *International Journal of Information and Electronics Engineering,* vol. 2, no. 5, pp. 678-681, ISSN: 2010-3719, 2012.

[6] Mohammad T. Kawser, Nafiz Imtiaz Bin Hamid, M. Nayeemul Hasan, M. Shah Alam, and M. Musfiqur Rahman, "Limiting HARQ Retransmissions in Downlink for Poor Radio Link in LTE", *International Journal of Information and Electronics Engineering,* vol. 2, no. 5, pp. 707-709, ISSN: 2010-3719, 2012.

[7] Available: http://new-techeurope.com/2016/02/01/lte-global-overview-and-lte-for-m2m-status/ [Accessed 2016]

[8] Kais Abdelrazeg El-Murtadi Suleiman, "Interactions Study of Self Optimizing Schemes in LTE Femtocell Networks", M. A. Sc. thesis, Queen's University, Canada, December 2012.

[9] Available: http://www.3glteinfo.com/lte-advanced-heterogeneous-networks/ [Accessed 2016]

[10] Available: https://frankrayal.com/2013/03/18/unleashing-the-power-of-hetnets-interference-management-techniques-for-lte-advanced-networks/ [Accessed 2016]

[11]   Sangmi Moon, Bora Kim, Saransh Malik, Daejin Kim, Youngil Kim, Kunmin Yeo, Intae Hwang, "Frequency and Time Resource Allocation for Enhanced Interference Management in a Heterogeneous Network based on the LTE-Advanced", The Ninth International Conference on Wireless and Mobile Communications (ICWMC), 2013.

[12]   Marwat, Safdar Nawaz Khan, et al. "Congestion-aware handover in LTE systems for load balancing in transport network." *ETRI Journal*, 36.5, pp: 761-771, 2014.

[13]   Jakub Bluszcz, "LTE measurement events", Leliwa Technical Bulletin, 2010.

[14]   3GPP Technical Specification 23.401, 'LTE; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access', 2014.

[15]   Available:http://ftp.actix.com/resources/1502/Projects/KB/MergedProjects/EventReference/GSM/GSM_Dragged_Handovers.htm [Accessed 2016]

[16]   Available: http://www.3glteinfo.com/lte-rlc-arq-procedure/ [Accessed 2016]

[17]   Available: http://www.dolcera.com/wiki/index.php?title=LTE [Accessed 2016]

[18]   3GPP Technical Specification 36.331, 'LTE; Radio Resource Control (RRC); Protocol specification', 2014.

[19]   3GPP Technical Specification 36.323, 'LTE; Packet Data Convergence Protocol (PDCP) specification', 2014.

[20]   3GPP Technical Specification 36.322, 'LTE; Radio Link Control (RLC) protocol specification', 2014.

[21]   3GPP Technical Specification 36.321, 'LTE; Medium Access Control (MAC) protocol specification', 2014.

[22]   Available: http://lteshare.blogspot.com/2013_12_01_archive.html [Accessed 2016]

[23]   3GPP Technical Specification 36.211, 'LTE; Physical channels and modulation', 2014.

[24]   3GPP Technical Specification 36.212, 'LTE; Multiplexing and channel coding', 2014.

[25]   3GPP Technical Specification 36.213, 'LTE; Physical layer procedures', 2014.

[26]   Zhou, Lei, Haibo Xu, Hui Tian, Youjun Gao, Lei Du, and Lan Chen. "Performance analysis of power saving mechanism with adjustable DRX cycles in 3GPP LTE." In Vehicular Technology Conference, (VTC 2008-Fall), IEEE 68th, pp. 1-5, 2008.

[27]   Mihov, Yakim Y., Kiril M. Kassev, and Boris P. Tsankov. "Analysis and performance evaluation of the DRX mechanism for power saving in LTE." In IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI), pp. 000520-000524, 2010.

[28]   Fowler, Scott, Ranjeet S. Bhamber, and Abdelhamid Mellouk. "Analysis of adjustable and fixed DRX mechanism for power saving in LTE/LTE-advanced." In IEEE International Conference on Communications (ICC), pp. 1964-1969, 2012.

[29]   Wang, Hwang-Cheng, Chih-Cheng Tseng, Guan-Yun Chen, Fang-Chang Kuo, and Kuo-Chang Ting. "Power saving by LTE DRX mechanism using a mixture of short and long cycles." In TENCON IEEE Region 10 Conference, pp. 1-6, 2013.

[30]   Fowler, Scott, George Baravdish, and Di Yuan. "Numerical analysis of an industrial power saving mechanism in LTE." In IEEE International Conference on Communications (ICC), pp. 1748-1753, 2014.

[31]   Jha, Satish Chandra, Ali Taha Koc, and Rath Vannithamby. "Optimization of discontinuous reception (DRX) for mobile Internet applications over LTE." In Vehicular Technology Conference (VTC Fall), pp. 1-5, IEEE, 2012.

[32]   Koc, Ali T., Satish Chandra Jha, Rath Vannithamby, and Murat Torlak. "Optimizing DRX configuration to improve battery power saving and latency of active mobile applications over LTE-A network." In Wireless Communications and Networking Conference (WCNC), pp. 568-573. IEEE, 2013.

[33]   Liu, Yu, Minh Huynh, Anurag Mangla, and Dipak Ghosal. "Performance analysis of adjustable discontinuous reception (DRX) mechanism in LTE network." In Wireless and Optical Communication Conference (WOCC), pp. 1-6. IEEE, 2014.

[34]   Rajesh, A., and R. Nakkeeran. "Performance analysis of enhanced DRX mechanism in LTE networks." In International Conference on Computer Communication and Informatics (ICCCI), pp. 1-5. IEEE, 2014.

[35]   Karthik, R. M., and Arvind Chakrapani. "Practical algorithm for power efficient drx configuration in next generation mobiles." In INFOCOM, pp. 1106-1114. IEEE, 2013.

[36]   Koc, Ali T., Satish C. Jha, Rath Vannithamby, and Murat Torlak. "Device power saving and latency optimization in LTE-A networks through DRX configuration." In IEEE Transactions on Wireless Communications, Vol. 13, Issue no. 5, pp. 2614-2625. 2014.

[37]   Wang, Kangping, Xin Li, Hong Ji, and Xin Du. "Modeling and Optimizing the LTE Discontinuous Reception Mechanism under Self-Similar Traffic.", In IEEE Transactions on Vehicular Technology, Vol. PP, Issue no.99, pp.1-1, 2013.

[38]   Y.-P. Yu and K.-T. Feng, "Traffic-Based DRX Cycles Adjustment Scheme for 3GPP LTE Systems," In IEEE Vehicular Technology Conference (VTC Spring), pp. 1–5, May 2012.

[39]   Yu, Gwo-Jong. "A fuzzy adaptive DRX power saving mechanism for LTE-Advanced networks." In The Fourth International Conference on Network of the Future (NOF), pp. 1-5, IEEE, 2013.

[40]   Syama Varma R, Krishna M. Sivalingam, Li-Ping Tung2 and Ying-Dar Lin,"Dynamic DRX Algorithms for Reduced Energy Consumption and Delay in LTE Networks" In IEEE Conference on Wireless Days (WD), pp. 1 – 8, 2014.

[41]   Zhang, Ziqi, Zhuyan Zhao, Hao Guan, Lei Du, and Zhenhui Tan. "Performance analysis of an adaptive DRX mechanism with flexible short/long cycle switching in LTE network." In IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE), pp. 27-32, 2013.

[42]   3GPP TS 23.203, "Policy and charging control architecture." Rel. 12, v. 12.9.0, 2014.

[43]   Moustafa M.Nasralla and Maria G. Martini, "A Downlink Scheduling Approach for Balancing QoS in LTE Wireless Networks" In Personal Indoor and Mobile Radio Communications (PIMRC), pp. 1571 - 1575, 2013.

[44]   Bontu, Chandra S., and Ed Illidge. "DRX mechanism for power saving in LTE." In Communications Magazine, IEEE, Vol. 47, Issue no. 6, pp. 48-55, 2009.

[45]   Jin, Sunggeun, and Daji Qiao. "Numerical analysis of the power saving in 3GPP LTE advanced wireless networks." , IEEE Transactions on Vehicular Technology, Vol. 61, Issue no. 4, pp. 1779-1785. IEEE, 2012.

[46]   Gupta, Pragya Kirti, R. V. Rajakumar, C. Senthil Kumar, and Goutam Das. "Analytical evaluation of signalling cost on power saving mechanism in mobile networks." In TENCON Spring Conference, pp. 376-380, IEEE, 2013.

[47]   Varma, Syama, Krishna M. Sivalingam, Li-Ping Tung, and Ying-Dar Lin. "Analytical model for power savings in LTE networks using DRX mechanism." In 2015 Twenty First National Conference on Communications (NCC), pp. 1-6, IEEE, 2015.

[48]   Fowler, Scott. "Study on power saving based on radio frame in LTE wireless communication system using DRX." In GLOBECOM Workshops (GC Workshops), pp. 1062-1066. IEEE, 2011

[49]   Wigard, Jeroen, Troels Kolding, Lars Dalsgaard, and Claudio Coletti. "On the user performance of LTE UE power savings schemes with discontinuous reception in LTE." In IEEE International Conference on Communications Workshops, 2009. ICC Workshops 2009, pp. 1-5. IEEE, 2009.

[50]   Zukerman, Moshe. "Introduction to Queuing Theory and Stochastic Teletraffic Models", http://www.ee city u.edu.hk/Zukerman/classnote pdf (2000-2015).

[51]   D. Lopez-Perez, I. Guvenc, and X. Chu, "Mobility management challenges in 3GPP heterogeneous networks," IEEE Commun. Mag., vol. 50, no. 12, pp. 70-78, 2012.

[52]   Meryem Simsek, Mehdi Bennis and Ismail Guvenc, "Mobility management in HetNets: a learning-based perspective", EURASIP Journal on Wireless Communications and Networking, 2015, DOI 10.1186/s13638-015-0244-2.

[53]   Technical Report 3GPP TR 36.839, "Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility Enhancements in Heterogeneous Networks", 2013.

[54]   Y Peng, YZW Yang, Y Zhu, "Mobility Performance Enhancements for LTE-Advanced Heterogeneous Networks", in Proc. IEEE Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)., Sydney, September, 2012.

[55]   M. P Wiley-Green, T Svensson, "Throughput, Capacity, Handover and Latency Performance in a 3GPP LTE FDD Field Trial", in Proc. IEEE GLOBECOM., FL, USA, December, 2010.

[56]   Fujitsu Network Communications, "Enhancing LTE Cell-Edge Performance via PDCCH ICIC", 2011.

[57]   K. Dimou, M. Wang, Y. Yang, M. Kazmi, A. Larmo, J. Pettersson, W. Muller, and Y. Timner, "Handover within 3GPP LTE: design principles and performance," in Proc. IEEE Vehic. Technol. Conf. (VTC), September 2009, pp. 1-5.

[58]   Qualcomm Inc., "LTE Advanced: Heterogeneous Networks", January 2011.

[59]   S. Barbera, P. Michaelsen, M. Saily, and K. Pedersen, "Mobility performance of LTE co-channel deployment of macro and pico cells," in Proc. IEEE Wireless Commun. and Networking Conf. (WCNC), pp. 2863-2868, April 2012.

[60]   X. Gelabert, G. Zhou, and P. Legg, "Mobility performance and suitability of macro cell power-off in LTE dense small cell HetNets," in Proc. IEEE Int. Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), pp. 99-103, September 2013.

[61]   D. Lopez-Perez, i. Guvenc, and X. Chu, "Mobility enhancements for heterogeneous wireless networks through interference coordination," in Proc. IEEE Int. Workshop on Broadband Femtocell Technologies (co-located with IEEE WCNC), Paris, France, Apr. 2012.

[62]   S. Barbera, P. Michaelsen, M. Saily, and K. Pedersen, "improved mobility performance in LTE co-channel HetNets through speed differentiated enhancements," in Proc. IEEE Globecom Workshops (GC Wkshps), pp. 426-430, Dec. 2012.

[63] Y. Kim, K. Lee, Y. Chin, "Analysis of Multi-level Threshold Handoff Algorithm," Proceeding of Global Telecommunication Conference, pp. 1141-1145, 1996.

[64] H. Schober and F. Jondral, "Velocity Estimation for OFDM Based Communication Systems," IEEE VTC-Fall 2002, vol. 2, pp. 715–718, Canada, 2002.

[65] T. Yucek, R. M. A. Tannious, and H. Arslan, "Doppler spread estimation for wireless OFDM systems," Proc. IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication, pp. 233–236, 2005.

[66] C. Tepedelenlioglu, A. Abdi, G. B. Giannakis, and M. Kaveh, "Estimation of Doppler spread and signal strength in mobile communications with applications to handoff and adaptive transmission," Wireless Communications and Mobile Computing, vol. 1, pp. 221-242, 2001.

[67] Partha Pratim Bhattacharya, "A New Environment Dependent Handoff Technique for Next Generation Mobile Systems," International Journal of Computer and Communications, Vol. 1, No. 1, 2011.

[68] Emrah Tunçel, "Tuning Of Handover Parameters In Lte-A Heterogeneous Networks", Middle East Technical University, M.Sc. Thesis, September, 2014.

[69] SHI, Y., MACKENZIE, A. B., DASILVA, L. A., et al. On resource reuse for cellular networks with femto and macrocell coexistence. In Proc. of the IEEE Global Telecommunication Conference (Globecom). Miami (FL, USA), Dec. 2010, 6 p. DOI: 10.1109/GLOCOM.2010.5683443

[70] FUJITSU NETWORK COMMUNICATIONS. 4G Femtocell for the Dense Metropolitan Environment, 2013. Available at: http://www.fujitsu.com/us/Images/4G-Femtocell-for-Dense-Metro-Environment.pdf

[71] NOVLAN, T., ANDREWS, J. G., ILLSOO SOHN, et al. Comparison of fractional frequency reuse approaches in the OFDMA cellular downlink. In Proc. of the IEEE Global Telecommunication Conference GLOBECOM 2010. Miami (FL, USA), Dec. 2010, 5 p. DOI: 10.1109/GLOCOM.2010.5683973

[72] SAQUIB, N., HOSSAIN, E., KIM, D. I. Fractional frequency reuse for interference management in LTE-advanced HetNets. IEEE Wireless Communications, April 2013, vol. 20, no. 2, p. 113–122. DOI: 10.1109/MWC.2013.6507402

[73] FRADI, N., NAJEH, S., BOUJEMAA, H. Resource allocation in OFDMA networks with femto and macro-cells coexistence using Fractional Frequency Reuse (FFR). In International Conference on Communications and Networking (ComNet). Hammamet (Tunisia), March, 2014, 5 p. DOI: 10.1109/COMNET.2014.6840931

[74]   NECKER, M. C. Local interference coordination in cellular OFDMA networks. In IEEE 66th Vehicular Technology Conference (VTC 2007 Fall). Baltimore (MD, USA), Sept. 2007, p. 1741–1746. DOI: 10.1109/VETECF.2007.368

[75]   LEI, H., ZHANG, L., ZHANG, X., et al. A novel multi-cell OFDMA system structure using fractional frequency reuse. In IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Athens (Greece), Sept. 2007, 5 p. DOI: 10.1109/PIMRC.2007.4394228

[76]   ASSAAD, M. Optimal Fractional Frequency Reuse (FFR) in multicellular OFDMA system. In IEEE Vehicular Technology Conference (VTC 2008 Fall). Calgary (BC), Sept. 2008, 5 p. DOI: 10.1109/VETECF.2008.381

[77]   HASSAN, N. U. L., ASSAAD, M. Optimal fractional frequency reuse (FFR) and resource allocation in multiuser OFDMA system. In Proceedings of International Conference on Information and Communication Technologies (ICICT 2009). Karachi (Pakistan), Aug. 2009, p. 88–92. DOI: 10.1109/ICICT.2009.5267207

[78]   3GPP, R1-050507, Huawei. Soft frequency reuse scheme for LTE, 2005.

[79]   LEE, P., LEE, T., JEONG, J. et al. Interference management in LTE femtocell systems using fractional frequency reuse. In Proc. of 12th International Conference on Advanced Communication Technology (ICACT). Phoenix Park, Feb. 2010, vol. 2, p. 1047 to 1051. ISBN: 978-1-4244-5427-3

[80]   CHANG, R. Y., TAO, Z., ZHANG, J., et al. A graph approach to dynamic fractional frequency reuse (FFR) in multi-cell OFDMA networks. In Proceedings of IEEE International Conference on Communications (ICC 2009). Dresden (Germany), June 2009, 6 p. DOI: 10.1109/ICC.2009.5198612

[81]   FANG, L., ZHANG, X. Optimal fractional frequency reuse in OFDMA based wireless networks. In Proceedings of 4th International Conference on Wireless Communications, Networking and Mobile Computing, (WiCOM'08). Dalian (China), 4 p. DOI: 10.1109/WICOM.2008.166

[82]   LEE, T., YOON, J., LEE, S., et al. Resource allocation analysis in OFDMA femtocells using fractional frequency reuse. In IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC). Istanbul (Turkey), 2010, p. 1224–1229. DOI: 10.1109/PIMRC.2010.5672042

[83] KUMAR, S., KOVACS, I. Z., MONGHAL, G., et al. Performance evaluation of a 6-sector-site deployment for downlink UTRAN long term evolution. In IEEE Vehicular Technology Conference (VTC 2008 Fall). Calgary (BC), Sept. 2008, 5 p. DOI: 10.1109/VETECF.2008.384

[84] SCANFERLA, D. Studies on 6-Sector-Site Deployment in Downlink LTE. Technische Universiteit Eindhoven (TUE), 2012. ISBN: 978-90-444-1128-7